

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

Путренко Віктор Валентинович

УДК 681.325:477

ДИСЕРТАЦІЯ
МЕТОДОЛОГІЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ГЕОПРОСТОРОВИХ
ДАНИХ ДЛЯ ЗАДАЧ СТАЛОГО РОЗВИТКУ

Спеціальність 01.05.04 – Системний аналіз і теорія оптимальних рішень

Подається на здобуття наукового ступеня доктора технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ В.В. Путренко

Науковий консультант
Згуровський Михайло Захарович
доктор технічних наук, професор,
академік НАН України

Київ – 2020

АНОТАЦІЯ

Путренко В.В. Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 01.05.04 – Системний аналіз і теорія оптимальних рішень Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, 2020.

Дисертаційна робота присвячена розробці методології інтелектуального аналізу геопросторових даних для задач сталого розвитку.

Підтримка прийняття управлінських рішень в територіально розподілених системах засновується на використанні геопросторової інформації, що в свою чергу вимагає застосування складних математичних та обчислювальних інструментів і методів інтелектуального аналізу та обробки геопросторових даних з метою досягнення цілей сталого розвитку та безпеки життя людей на відповідних територіях.

Зазначені тенденції дають поштовх для розвитку індустрії геопросторових даних, переходу до управління територіями (велике місто, регіон країни, країна) на основі використання геоінформаційних систем, створення нових систем та методів збору геопросторових даних, в тому числі геокодування та геотегування інформації, накопичення великих обсягів геопросторових даних. Ці задачі вимагають застосування новітніх способів аналізу великих обсягів геопросторових даних з метою виявлення прихованих закономірностей в їх структурі. Тому інтелектуальний аналіз геопросторових даних, як важливий напрям наукових досліджень, відкриває нові можливості для побудови ефективних систем територіального управління.

В дисертаційній роботі розв'язано важливу науково-прикладну проблему інтелектуального аналізу геопросторових даних з метою розпізнавання прихованих закономірностей та відношень в задачах сталого розвитку територіально розподілених систем.

Розроблено методологію та обґрунтовано доцільність використання інтелектуальних методів аналізу геопросторових даних щодо сталого розвитку на засадах методів системного аналізу.

Розроблено теоретико-методологічні підходи до формалізації поняття та моделей представлення геопросторових даних на основі парадигми дискретних та континуальних ознак тривимірного простору та його часової зміни.

Визначено, що базовими моделями організації геопросторових даних є векторні та растрові моделі.

Формалізація моделей геопросторових об'єктів потребує визначення системи координат та її уніфікації для сумісного використання геоданих в геоінформаційній системі. Зокрема розглянуто особливості застосування і сумісного використання найбільш розповсюджених в Україні географічної системи координат.

Набори геопросторових даних формуються та структуруються у рамках інфраструктури геопросторових даних (ІПД) - ієрархічної впорядкованої системи, яка побудована з використанням кібертехнологій та геоінформаційних підсистем, заснована на загальних просторових даних і метаданих, а також мережі географічних інформаційних вузлів - геопорталов і каталогів метаданих.

Основною метою використання інтелектуального аналізу геопросторових даних (ІАГД) є пошук закономірностей та відносин у великих масивах даних, які містять просторово координовану прив'язку. Тому використання ІАГД в складі СППР в територіальному управлінні та прогнозуванні є важливим та актуальним інструментом обґрунтування управлінських рішень. Процес реалізації ІАГД відрізняється стадійністю та ітераційністю з метою пошуку оптимальних результатів аналізу. До основних напрямів ІАГД відносяться класифікація, кластеризація, правила асоціацій, геостатистика та геовізуалізація, які разом формують методологію системного інтелектуального аналізу з метою підтримки прийняття рішень. ІАГД підвищує ефективність обробки даних разом з іншими методами аналізу даних на основі різних інформаційних платформ.

Запропоновано системний підхід до методів інтелектуального аналізу даних на основі геопросторових даних. Визначено ключові параметричні особливості цих

методів, які полягають у геопросторовій координатності, топологічній єдності, геопросторових зв'язках та відносинах між собою. В роботі розв'язано шість базових видів задач використання методів ІАГД для аналізу метрики сталого розвитку.

Запропоновано для вирішення шість типів задач удосконалення метрики сталого розвитку, в тому числі:

1. Визначення впливу просторових відносин на показники сталого розвитку
2. Відновлення пропусків значень у статистичних рядах метрики сталого розвитку
3. Виділення однорідних територіальних утворень для реалізації заходів регіональної політики
4. Аналіз безпекового виміру на різних територіальних рівнях
5. Аналіз екологічної компоненти землекористування
6. Розрахунок метрики сталого розвитку на основі даних ДЗЗ

Базові методи інтелектуального аналізу геопросторових даних визначаються типом розподілу даних та гіпотезою імовірності оцінки виникнення аномальних значень на обмеженій просторовій відстані. Для визначення цих параметрів використовується Просторова автокореляція між геопросторовими об'єктами (Глобальний індекс І Морана), показник середнього найближчого сусідства, Локальний індекс Морана. Для цілей просторової кластеризації застосовується аналіз гарячих точок (Getis-Ord G_i^*), ефективність угруповання об'єктів вимірюється за допомогою псевдо-F-статистики Калінскі-Харабаза. Просторова кластеризація визначається просторовими обмеженнями топології. Важливою складовою інтелектуального аналізу геопросторових даних є моделювання просторових відносин методами зворотної відстані, діапазони відстаней, зон індиферентності, суміжності та сусідства.

Розроблено підхід до аналізу великих геопросторових даних шляхом їх дворівневого аналізу за допомогою організації даних у просторово- часових кубах, де на основі методів просторової кластеризації відбувається виділення

інформаційних паттернів даних. Апробація методики здійснена на базі даних про події у телекомунікаційній мережі оператора мобільного зв'язку Vodafone, що дозволило вирішити задачу аналізу безпекового виміру на різних територіальних рівнях.

Застосування методів класифікації та кластерного аналізу у поєднанні з аналітичними можливостями геоінформаційних систем значно розширює методичний інструментарій сучасного інтелектуального аналізу даних.

На основі канонічної системи методів класифікації сформована система методів районування. Розібрані алгоритми деяких найбільш застосовних на практиці методів районування, які є модифікаціями канонічних методів, і метод бар'єрів максимальних відмінностей. Розроблено алгоритм побудови моделі сукупності однорідних територіальних зон. Створені засоби проведення нечіткого кластерного аналізу у середовищі ArcGIS дозволяють класифікувати просторові об'єкти за просторово-атрибутивними параметрами. До основних переваг багатовимірного аналізу в середовищі ГІС слід віднести автоматизоване картографування кластерів, використання додаткової географічної інформації, варіабельність у прийнятті рішень щодо аналізу. Було здійснено апробацію розробленого підходу на основі кластерного аналізу даних про параметри сталого розвитку в Україні.

Застосовано методи геостатистичного аналізу даних для аналізу параметрів сталого розвитку України з використанням детермінованих та стохастичних методів аналізу даних, які включають процедури побудови глобальних інтерполяційних моделей на основі трендів, методу обернено зваженої відстані, сплайнів та методології крігінгу. В результаті доведено доцільність застосування трендових поверхонь для визначення глобальних трендів регіональної зміни параметрів сталого розвитку на прикладі регіонів України, що може бути застосовано у моделях передбачення та прогнозування.

На основі даних про вміст миш'яку у підземних водах на території України було досліджено детерміновані та стохастичні методи інтерполяції. В результаті було удосконалено застосування методу крігінгу на основі підбору сферичної

моделі варіограми, що дозволило отримати найкращі статистичні оцінки точності для інтерпольованої поверхні методом сферичного крігінгу.

Розроблено методику навчання та використання багатошарових згорткових нейронних мереж для визначення параметрів сталого розвитку території незалежно від адміністративно-територіального поділу території на прикладі аналізу індексу економічного виміру території.

Було застосовано навчання згорткової нейронної мережі з градієнтним спуском на основі мультиспектральних зображень та знімків нічного освітлення, які дозволили з високою надійністю розпізнавати на знімках території міських та сільських населених пунктів, промислових зон та сільськогосподарських угідь. Під час навчання було застосовано різні варіанти функцій активації згорткової мережі. Найкращий результат було досягнуто з використанням функції softmax. Надійність навчання мережі склала 87%. Це дозволило вперше оцінити значення індексу виміру економічного розвитку для окупованих територій України та АР Крим, сформувати розподіл значень індексу для адміністративних одиниць рівня району та міста, а також визначити динаміку індексу за останні роки з використанням інформації з космічних знімків.

В результаті можна зазначити, що використання методу згорткових нейронних мереж може дозволити отримати кількісні дані, що характеризують процес сталого розвитку території незалежно від статистичних джерел інформації з високим ступенем надійності. Основними властивостями цих даних є розрізнення вхідних космічних даних, крок мережі для апроксимації інформації та метод навчання нейронної мережі.

Наукова новизна роботи визначається наступними теоретичними і практичними результатами, отриманими автором:

Наукова новизна: У дисертації сформульовано та обґрунтовано низку наукових положень, які у своїй сукупності дозволяють розв'язати важливу науково-прикладну проблему інтелектуального аналізу геопросторових даних з метою розпізнавання прихованих закономірностей та відношень в структурі цих даних при вирішенні задач сталого розвитку великих територіально розподілених систем.

Основні наукові результати дисертаційної роботи полягають у наступному:

вперше:

- розроблено методологію використання інтелектуальних методів аналізу геопросторових даних, яка включає в себе принципи сталого розвитку, моделі, методи та технології системного управління територіями;
- розроблено методику використання геопросторових даних дистанційного зондування Землі для отримання інтегральної оцінки параметрів сталого розвитку та безпеки життя людей:
 - розроблено метод моделювання на основі географічно зваженої регресії та локального індексу Морана;
 - розроблено метод розрахунку індексу економічного розвитку території з використанням нейронної мережі на основі навчання на базі даних дистанційного зондування Землі.
- розроблено методику трендового аналізу функціонального та кількісного зв'язку вкладу складових індексу сталого розвитку
- розроблено методику районування території за параметрами сталого розвитку з використанням топологічних геопросторових даних місцеположення з використанням методів геопросторової класифікації та кластеризації

удосконалено:

- компонент безпеки моделі сталого розвитку на основі індексу природних та техногенних небезпек.
- методи геостатистичного аналізу та оптимізації задачі інтерполяції даних стохастичними методами крігінгу.
- методи аналізу просторово-часого кубу даних у дослідженні великих геопросторових даних.

отримали подальший розвиток:

- методи просторової кластеризації та нечіткої кластеризації даних з використанням статистичної обробки «гарячих-холодних» точок
- методи геопросторової концептуалізації топологічних відносин між об'єктами
- методи глобальної та локальної просторової регресії для вирішення завдань оптимізації конфігурування просторових паттернів у розподілі точкових об'єктів.

Практичне значення одержаних результатів полягає у створенні інструментарію СППР на основі програмної платформи ArcGIS для інтеграції методів інтелектуального аналізу та нейронних мереж до геоінформаційної системи з метою оцінки параметрів сталого розвитку території з позицій кластеризації, районування та комплексування адміністративно-територіальних одиниць для досягнення цілей сталого розвитку та безпеки життя людей. Запропонований підхід до застосування нейронних мереж дозволяє оцінити параметри та інтегральну оцінку складових сталого розвитку на різних територіальних рівнях, які відповідають європейській класифікації регіонів NUTS 1-5 та державного адміністративно-територіального устрою України, включаючи новий тип адміністративних територіальних утворень – об'єднані територіальні громади.

Другим аспектом застосування інструментарію є можливість отримання необхідних даних для територій, на яких статистичні дані не збираються. В Україні до класу таких територій відноситься окупована частина Донбасу та АР Крим.

Результати дисертаційної роботи впроваджено в дослідженнях та звітах ННК «Світовий центр даних з геоінформатики та сталого розвитку» в межах дослідницьких робіт та ініціативних тематик аналізу параметрів сталого розвитку країн світу та України.

Результати дисертаційної роботи впроваджено в навчальний процес кафедри математичних методів системного аналізу Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний

інститут імені Ігоря Сікорського», зокрема при викладанні дисципліни «Геоінформаційні системи» та «Інтелектуальний аналіз даних в ГІС».

Ключові слова: інтелектуальний аналіз, геопросторові дані, геоінформаційна система, кластеризація, нейронні мережі, трендовий аналіз, геостатистика, сталий розвиток, якість життя.

SUMMARY

Putrenko V.V. Methodology for the intellectual analysis of geospatial data for sustainable development goals. - Qualifying scientific work on the rights of the manuscript.

Thesis for a Doctor of Science Degree in Specialty 01.05.04 - System Analysis and Theory of Optimal Solutions, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 2020.

The dissertation is devoted to the development of methodology for the intellectual analysis of geospatial data for sustainable development goals.

Supporting management decision-making in the management of geographically distributed systems is based on the use of geospatial information, which in turn requires the use of techniques and methods of intellectual analysis and processing of geospatial data to achieve the goals of sustainable development. The main reasons for this are the development of the geospatial data industry, the transition to management based on the use of geoinformation systems, the creation of new systems and methods of geospatial data collection, including geocoding and geotagging of information, the accumulation of large amounts of geospatial data, which require new ways of analyzing and finding patterns in their structure. Therefore, data mining opens new opportunities for finding optimal management solutions at all levels of territorial governance.

The dissertation deals with the important scientific and applied problem of the intellectual analysis of geospatial data in order to recognize the hidden patterns and relations in the problems of sustainable development of territorially distributed systems.

The methodology has been developed and the expediency of using intelligent methods of geospatial data analysis for sustainable development based on the methods of system analysis has been substantiated.

Theoretical and methodological approaches to formalization of the concept and models of representation of geospatial data based on the paradigm of discrete and continuous features of three-dimensional space and its temporal change are developed.

It is determined that the basic models of organization of geospatial data are vector and raster models.

Formalizing geospatial object models requires defining a coordinate system and unifying it to share location data in a geoinformation system. In particular, the features of application and sharing of the most common geographic coordinate system in Ukraine are considered.

Geospatial data sets are formed and structured within the Geospatial Data Infrastructure (GDI) - a hierarchical orderly system that is built using cyber technologies and geoinformation subsystems, based on shared spatial data and metadata, as well as a network of geographic information nodes - geoportals.

The primary purpose of using Geospatial Data Intelligence (GDI) is to search for patterns and relationships in large datasets that contain spatially coordinated binding. Therefore, the use of GDI as part of the DSS in territorial management and forecasting is an important and relevant tool for substantiating management decisions. The implementation process of GDI is stable and iterative in order to find optimal analysis results. The main directions of GDI are classification, clustering, rules of associations, geostatistics and geo-visualization, which together form the methodology of systematic intellectual analysis to support decision-making. GDI improves data processing efficiency with other data analysis methods based on different information platforms.

A systematic approach to data mining methods based on geospatial data is proposed. The key parametric features of these methods are defined, which are geospatial coordination, topological unity, geospatial connections and relationships. The work addresses six basic problems of using GDI methods for analyzing sustainable development metrics.

It is proposed to solve six types of sustainable development metric improvement tasks, including:

1. Determining the influence of spatial relations on indicators of sustainable development.
2. Restoration of value gaps in statistical series of metrics for sustainable development.
3. Allocation of homogeneous territorial entities for implementation of regional policy measures.

4. Analysis of safety measurement at different territorial levels.
5. Analysis of the environmental component of land use.
6. Sustainability metric calculation based on remote sensing data.

Basic methods of geospatial data mining are determined by the type of data distribution and the hypothesis of the probability of estimating the occurrence of anomalous values over a limited spatial distance. The spatial autocorrelation between geospatial objects (Global I Moran Index), the mean nearest neighbor index, the Moran Local Index is used to determine these parameters. For the purposes of spatial clustering, hotspot analysis (Getis-Ord Gi) is used, and the grouping of objects is measured using the Kalinski-Kharabaz pseudo-F statistics. Spatial clustering is determined by the spatial constraints of the topology. An important component of the intellectual analysis of geospatial data is the modeling of spatial relationships by back-distance methods, ranges of distances, zones of indifference, adjacency and neighborhood.

An approach to the analysis of large geospatial data by their two-level analysis with the help of data organization in space-time cubes, where based on methods of spatial clustering, the allocation of information patterns of data is developed. The methodology was tested on a database of events on the telecommunication network of the Vodafone mobile operator, which made it possible to solve the problem of security measurement analysis at different territorial levels.

The application of the methods of classification and cluster analysis in combination with the analytical capabilities of geoinformation systems greatly expands the methodological tools of modern data mining.

On the basis of the canonical system of classification methods, a system of zoning methods was formed. The algorithms of some of the most applicable zoning methods, which are modifications of the canonical methods, and the method of barriers of maximum differences, are analyzed. An algorithm for constructing a model set of homogeneous territorial zones has been developed. The fuzzy cluster analysis tools created in ArcGIS allow you to classify spatial objects by their spatial attribute parameters. The main advantages of multidimensional analysis in a GIS environment include the automated mapping of clusters, the use of additional geographical

information, and variability in decision-making on analysis. The developed approach was tested on the basis of cluster analysis of data on sustainable development parameters in Ukraine.

Methods of geostatistical data analysis were applied to analyze the parameters of sustainable development of Ukraine using deterministic and stochastic data analysis methods, which include procedures for building global interpolation models based on trends, inverse weighted distance method, splines and kriging methodology. As a result, the expediency of using trend surfaces to determine the global trends of regional change in the parameters of sustainable development on the example of regions of Ukraine is proved, which can be applied in models of forecasting and forecasting.

Deterministic and stochastic interpolation methods were investigated on the basis of arsenic content in groundwater in Ukraine. As a result, the application of the kriging method was improved based on the selection of the spherical model of the variogram, which allowed to obtain the best statistical estimates of accuracy for the interpolated surface by the method of spherical kriging.

The methodology of training and use of multilayer backbone neural networks to determine the parameters of sustainable development of the territory, regardless of the administrative-territorial division of the territory on the example of the analysis of the economic dimension of the territory.

A convolutional neural network training with gradient descent based on multispectral images and night-light images was applied, which made it possible to recognize with high reliability the images of urban and rural settlements, industrial zones and agricultural land. Various variants of the convolution network activation function were applied during the training. The best result was achieved using the softmax function. Network learning reliability was 87%. It allowed for the first time to estimate the value of the index of measurement of economic development for the occupied territories of Ukraine and Crimea, to form the distribution of index values for the administrative units of the district and city level, as well as to determine the dynamics of the index in recent years using information from space images.

As a result, it can be noted that the use of the convolutional neural network method can allow to obtain quantitative data characterizing the process of sustainable development of the territory, regardless of the statistical sources of information with a high degree of reliability. The main properties of this data are distinguishing the input space data, the network step to approximate information, and the neural network training method.

The scientific novelty of the work is determined by the following theoretical and practical results obtained by the author:

Scientific novelty: The thesis defines and substantiates a number of scientific provisions, which in their essence solve the important scientific and applied problem of the problem of intellectual analysis of geospatial data in order to recognize the hidden patterns and relations in the problems of sustainable development of territorially distributed systems.

The main scientific results of the dissertation are as follows:

for the first time:

- methodology for the use of intelligent methods of geospatial data analysis, which includes the principles of sustainable development, models, methods and technologies of systematic management of territories was developed;

The method of using geospatial data of remote sensing of the Earth to obtain a synthetic assessment of the parameters of sustainable development and safety of human life was developed:

- modeling method based on geographically weighted regression and local Moran index was developed;
- the method for calculating the index of economic development of the territory using a neural network based on training on the basis of remote sensing Earth.
- the methodology of trend analysis of functional and quantitative linkage of contribution of components of sustainable development index has been developed

- the methodology for zoning the territory according to the parameters of sustainable development using topological geospatial location data using methods of geospatial classification and clustering was developed

improved:

- the safety component of the sustainable development model based on the index of natural and man-made hazards.
 - methods of geostatistical analysis and optimization of the problem of data interpolation by stochastic kriging methods.
 - methods for analyzing the space-time cube of data in the study of large geospatial data.

were further developed:

- methods of spatial clustering and fuzzy data clustering using statistical hot-cold point processing
- methods of geospatial conceptualization of topological relations between objects
- global and local spatial regression methods to solve the problems of optimizing the configuration of spatial patterns in the distribution of point objects.

The practical significance of the obtained results is to create a DSS tool based on the ArcGIS software platform for integrating methods of intellectual analysis and neural networks into the geoinformation system in order to evaluate the parameters of sustainable development of the territory from the standpoint of clustering, zoning and complexation of administrative and territorial units for the development of security units people's life. The proposed approach to the use of neural networks allows to evaluate the parameters and integrated assessment of components of sustainable development at different territorial levels, which correspond to the European classification of regions NUTS 1-5 and state administrative-territorial structure of Ukraine, including a new type of administrative territorial entities - unified territorial entities.

The second aspect of the toolkit is the ability to obtain the necessary data for territories where statistics are not collected. In Ukraine, the occupied part of the Donbass and Crimea belongs to the class of such territories.

The results of the dissertation were implemented in the research and reports of NNC "World Data Center for Geoinformatics and Sustainable Development" within the framework of research works and initiative topics of analysis of the parameters of sustainable development of countries of the world and Ukraine.

The results of the dissertation were introduced into the educational process of the Department of Mathematical Methods of System Analysis of the Institute of Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", in particular in teaching the discipline "Geoinformation Systems" and "Intelligent Data Analysis in GIS".

Keywords: intellectual analysis, geospatial data, geoinformation system, clustering, neural networks, trend analysis, geostatistics, sustainable development, quality of life.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. **Путренко В.В.** Ресурсозбереження: регіональний аспект. - Наук. ред. Лісовський С.А. – К.: Інститут географії НАН України, 2011. – 224 с.
2. Геоінформаційне картографування в Україні. Концептуальні основи і напрями розвитку. – За ред. акад. Л.Г. Руденка. – К., Науково-виробниче підприємство «Видавництво «Наукова думка» НАН України», 2011. – 103 с.
3. Форсайт економіки України: середньостроковий (2015–2020 роки) і довгостроковий (2020–2030 роки) часові горизонти (версія для обговорення) / наук. ке- рівник проекту акад. НАН України М. З. Згуровський // Міжнародна рада з науки (ICSU); Національний технічний університет України «Київський політехнічний інсти- тут»; Інститут прикладного системного аналізу НАН України і МОН України; Світовий центр даних з геоінформатики та сталого розвитку. – Київ : НТУУ «КПІ», 2015. – 36 с.
4. Форсайт та побудова стратегії соціально-економічного розвитку України на середньостроковому (до 2020 року) і довгостроковому (до 2030 року) часових горизонтах / наук. керівник проекту акад. НАН України М. З. Згуровський // Міжнародна рада з науки (ICSU); Комітет із системного аналізу при Президії НАН України; Національний технічний університет України «Київський політехнічний інститут»; Інститут прикладного системного аналізу МОН України і НАН України; Світовий центр даних з геоінформатики та сталого розвитку. — Київ : НТУУ «КПІ», 2016. — 180 с.
5. Foresight of Ukrainian Economy: mid-term (2015–2020) and long-term (2020–2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky // International Council for Science (ICSU) ; Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Kyiv Polytechnic Institute» ; Institute for Applied System Analysis of NAS of Ukraine and MES of Ukraine ; World Data Center for Geoinformatics and Sustainable Development. — 2nd ed. — Kyiv : NTUU «KPI», Publ. house «Polytechnica», 2016. — 136 p.

6. Аналіз соціально-економічних процесів розвитку суспільства, заснованого на знаннях / наук. кер. проекту М. З. Згуровський. – К.: КПІ ім. Ігоря Сікорського, 2017. – 26 с. ISBN 978-966-622-819-5
7. Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for Science etc.; Scientific Supervisor M. Zgurovsky. – К. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. 226 p
8. Аналіз сталого розвитку – глобальний і регіональний контексти / Міжнар. рада з науки (ICSU) та ін.; наук. кер. проекту М. З. Згуровський. – К. : НТУУ «КПІ», 2013. – Ч. 2. Україна в індикаторах сталого розвитку (2013). – 178 с.
9. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p.
10. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — К. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016– 2017). — 72 p.
11. Govorov M., **Putrenko V.**, Gienko G. Mining Spatial Patterns of Distribution of Uranium in Surface and Ground Waters in Ukraine / Sami Faiz, Khaoula Mahmoudi. Handbook of Research on Geographic Information Systems Applications and Advancements, 2016, 567 p., що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
12. **Путренко В. В.** Перспективи розвитку Національного атласу України в мультимедійному напрямі // Вісник геодезії та картографії – 2009. – № 6. – С. 14 – 20. що входить до наукометричних баз Google Scholar, РИНЦ та ін.
13. **Путренко В. В.** Геоінформаційне картографування техногенних ризиків на території України: формування баз даних // Вісник геодезії та картографії – 2010. – № 5. – С. 23 – 28, що входить до наукометричних баз Google Scholar, РИНЦ та ін.

14. **Путренко В. В.** Картографічний підхід до оцінювання техногенних ризиків виникнення надзвичайних ситуацій на регіональному рівні // Вісник геодезії та картографії. – 2012. - №6. – С. 24-32, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
15. **Путренко В.В.** Системные основы интеллектуального анализа геопространственных данных / В.В. Путренко // Системні дослідження та інформаційні технології. – 2015. – № 3 - С. 24 – 32, що входить до наукометричних баз DOAJ, Index Copernicus, Google Scholar, РИНЦ та ін.
16. **Путренко В.В.,** Тихоход В.О. Розробка геостатистичних моделей інтелектуального аналізу забруднення повітря // Східно-Європейський журнал передових технологій. – 2015. - №4/10. – С. 21-26, що входить до наукометричних баз DOAJ, що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
17. **Путренко В.В.,** Красовська І.Г. Кластеризація геопросторових даних при інтелектуальному аналізі // Радіoeлектронні і комп'ютерні системи. – ХАІ. – 2015. - № 3. – С. 32-41, DOAJ, що входить до наукометричних баз Index Copernicus, Google Scholar, РИНЦ та ін.
18. **Путренко В.В.** Застосування трендових геопросторових статистичних моделей на основі інтелектуальному аналізі даних/ В.В. Путренко // Системні дослідження та інформаційні технології. – 2015. – № 4. – С7 – 19, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
19. **Путренко В.В.** Інтелектуальний аналіз небезпеки виникнення природних пожеж на основі геоінформаційних технологій // Технологический аудит и резервы производства. – Х., Технологический центр. – 2016. – С. 67 – 72, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
20. **V. Snytyuk, N. Pashynska, V. Putrenko, A. Musienko** A decision tree in a classification of fire hazard factors // Eastern-European Journal of Enterprise Technologies, 2016. - Vol 5, No 10(83) - P. 32-38, що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.

21. **Путренко В.В.** Застосування інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки / В. В. Путренко, Н. М. Пашинська // Вісник НТУ «ХПІ», Серія: Нові рішення в сучасних технологіях. – Харків: НТУ «ХПІ». – 2017. – 7 (1229). – С. 156-163. – doi:10.20998/2413-4295.2017.07.22, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
22. **Путренко В.В.,** Пашинська Н.М. Інтелектуальний аналіз геопросторових даних як інструмент планування сталого розвитку // Вісник Київського національного університету імені Тараса Шевченка. Серія фізико-математичні науки. 2017, Вип 1. С. 49-55, що входить до наукометричних баз DOAJ, Index Copernicus, Google Scholar, РИНЦ та ін.
23. **Putrenko V.** Pashynska N. The use of remote sensing data for modeling air quality in the cities // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-5/W1, 2017, pp. 57-62 що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
24. **Путренко В.В.,** Пархоменко А.В. Геопросторовий аналіз та візуалізація даних про військову конфліктогенність // Містобудування та територіальне планування: Наук.-техн. збірник / Головн. ред. М.М. Осетрін. – К., КНУБА, 2017. – Вип. 64. – С. 326 – 332, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
25. **Путренко В.В.,** Назаренко С.Ю., Пашинська Н.М. Моделювання даних з просторово-часовим розподілом для потреб прийняття рішень // Містобудування та територіальне планування: Наук.-техн. збірник / Головн. ред. М.М. Осетрін. – К., КНУБА, 2018. – Вип. 68. – С. 750 – 760, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
26. **Путренко В.В.,** Назаренко С.Ю. Визначення якості повітря на основі інтелектуального аналізу даних дистанційного зондування // Математичне моделювання в економіці, 2016. – No 3-4 (7). – С. 176-188, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.

27. **Путренко В.В.**, Пашинська Н.М. Інтелектуальний аналіз земного покриття території України на основі даних Globeland 30 // Управління розвитком складних систем. - Вип 31. 2017. – С. 117-126, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
28. Даценко Л.М., Говоров М., **Путренко В.В.**, Пашинська Н.М. Перші результати реалізації українсько-канадського освітнього проекту з питань закладення основ інфраструктури просторових даних // Вісник геодезії та картографії, Вип.4, 2015. – С. 46-52, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
29. **Путренко В.В.**, Бенатов Д. Е., Стефанишин Д . В . Геоінформаційна система «гідровузли України» – важливий елемент підтримки управлінських процедур // Восточно-Европейский журнал передовых технологий. - 1/3 (79) 2016. – 46 – 53, що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
30. **Putrenko V.**, Bowkun Alexandra Spatial modeling of the risk of accident in the power system // Risk information management, risk models and applications, 2015. – CODATA-Germany LNIS. – Vol.7 - P.1 – 13, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
31. **Putrenko V.**, Pashynska N. Risk Modeling of Accidents in the Power System of Ukraine with Using Bayesian Network // Advances in Computer Science for Engineering and Education, Springer International Publishing AG, 2018. – vol. 754. pp.13-22, що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
32. **Putrenko V.** Data mining of the land cover characteristics of Ukraine using the data of remote sensing of Earth // Slovak international scientific journal, VOL.1, №30, 2019. – р. 13-18. що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
33. Govorov M., Beconyè G., Gienko G., **Putrenko V.** Spatially constrained regionalization with multilayer perceptron // Transactions in GIS. 2019 Vol 23,

- Issue 4, p.1–30. що входить до наукометричних баз SCOPUS, Index Copernicus, Google Scholar та ін.
- 34.**Putrenko V.**, Pashynska N. Data mining of sustainable development process with using nightlight indicators // Technology audit and production reserves — № 3/2(47), 2019. — р. 4 – 9, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
35. **Путренко В.В.**, Пашинська Н.М. Регіональний моніторинг стану рослинності з використанням методів інтелектуального аналізу даних // Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки. - Том 30 (69) Ч. 1 № 4 2019. – С. 112 – 118, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
- 36.Комп'ютерна програма «Геоінформаційна система «Гідровузли України»:свідoctво про реєстрацію авторського права на твір 64745 Україна / Бенатов Д. Е., **Путренко В.В.**, Стефанишин Д. В., Розщупкін Г. В. – № 65227; заявл. 09.02.16; реєстр. 01.04.16.
- 37.Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том1: Вступ до геоінформаційних систем для інфраструктури просторових даних. Навчальний посібник / Б. Магваїр, М. Говоров, Н.М. Пашинська, Л.М.Даценко, **В.В. Путренко**. – Київ, 2016. – 396 с.
- 38.Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том 2: Системи керування базами геоданих для інфраструктури просторових даних. Навчальний посібник. / Д. Кейк, А.А. Лященко, **В.В. Путренко**, Ю. Хмелевський, К.С. Дорошенко, М. Говоров – Київ, 2017. – 456 с.
- 39.Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том 3: Просторові кадастрові інформаційні системи для інфраструктури просторових даних. Навчальний посібник. / М. Говоров, А.А. Лященко, Д. Кейк, П. Зандберген, М.А. Молочко, Л. Бевайніс, Л.М. Даценко, **В.В. Путренко** – Планета-Прінт, 2017. – 532 с.

- 40.**Путренко В.В.** Інтелектуальний аналіз земного покриття території України на основі даних Globeland 30 // Вісник Астрономічної школи, 2016, том 12, № 1 – С. 46 – 57, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
- 41.Zgurovsky M **Putrenko V.**, Pashynska N. Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine // 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), 2018. – pp. 8-12, що входить до наукометричних баз **SCOPUS**, ISI.
- 42.**Putrenko V.** Data Mining of Relationship in Crowdsourcing Projects and Social Activities of Citizens // 2017 IEEE First International Conference on electrical and computer engineering (UKRCON), conference proceedings, Kyiv, Ukraine, May 29 – June 2, 2017. – P. 1060 – 1066, що входить до наукометричних баз **SCOPUS**, ISI.
- 43.**Putrenko V.**, Pashynska N. Data Mining of Network Events with Space-Time Cube Application // IEEE Second International Conference on Data Stream Mining & Processing, August 21-25, 2018, Lviv, Ukraine. – pp. 79-83, що входить до наукометричних баз **SCOPUS**, Index Copernicus, Google Scholar та ін.
- 44.Chabaniuk, V.S., **Putrenko, V.V.** Persönlicher Atlas des Ministers für Katastrophenschutz der Ukraine // Diskussionsbeiträge zur Kartosemiotik und zur Theorie der Kartographie Intern. Korrespondenz-Seminar, Band 15. Dresden 2012, p. 14- 19, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
- 45.**Putrenko V.**, Pashynska N., Nazarenko S. Data Mining of Network Events With Space-Time Cube Application // First Workshop on Platial Analysis (PLATIAL'18), Heidelberg, 20–21 September 2018. – pp. 73-80, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
- 46.**Putrenko V.**, Modeling Spatial Relationships of Electoral Behavior in Ukraine // European review of law and international relations Chief of the Board Jerzy J. Wiatr, Warszawa, 2018. – P. 46 – 52, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.

47. **Путренко В.В.,** Пашинська Н. М., Назаренко С. Ю. Картографування якості повітря на основі аналізу даних дистанційного зондування // Часопис картографії: Збірник наукових праць. – К. : КНУ ім. Тараса Шевченка, 2016. – Вип. 15. Частина 1. – 89-98, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
48. **Путренко В. В.** Визначення концентрації оксиду азоту в атмосферному повітрі за даними дистанційного зондування Землі / В.В. Путренко, Н.А. Куцина, С.Ю. Назаренко // Праці Центральної геофізичної обсерваторії. – К.: Інтерпрес ЛТД, 2017. – Вип. 13 (27) – С. 55 – 61, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
49. **Putrenko V.,** Govorov M., Gienko G. Exploring Distribution of Uranium in Ukraine: Geovisualization and Spatial Statistics // 26th International Cartographic Conference, Dresden August 25 – 30, 2013 Proceedings. Режим доступу: <http://www.icc2013.org/?node=29>, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
50. **Putrenko V.** Assessment of Risk and Hazard of Technological Emergencies with GIS (on the Example Vinnytsia Region) // 26th International Cartographic Conference, Dresden August 25 – 30, 2013 Proceedings. Режим доступу: <http://www.icc2013.org/?node=29>, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
51. **Putrenko V.,** Govorov M., Gienko G. Exploring Spatial Patterns of Uranium Distribution in Ukraine // 16th AGILE Conference on Geographic Information Science: Geographic Information Science at the Heart of Europe: 14-17 May 2013, Leuven, Belgium. Режим доступу: <http://agile.gis.geo.tu-dresden.de/web/index.php/conference/proceedings>, що входить до наукометричних баз Index Copernicus, Google Scholar та ін.
52. **Путренко В.В.,** Тихоход В.А. Використання кластерного аналізу для просторового моделювання в геоінформаційних системах // Географічна наука і практика: виклики епохи: Матеріали міжнародної наукової конференції, присвяченої 130-річчю географії у Львівському університеті (м. Львів, 16 –

18 травня 2013 р.). У 3-ох томах. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2013. – Том 3. – 306 с., що входить до наукометричних баз Google Scholar, РИНЦ та ін.

53. **Путренко В.В.**, Мультимасштабні картографічні моделі аналізу небезпек та ризиків надзвичайних ситуацій // Ученые записки Таврического национального университета имени В.И.Вернадского Серия «География». Том 26 (65). 2013 г. № 1, С. 136-144, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
54. **Путренко В.В.**, Чабанюк В.С., Станкевич Т.В. Питання публікації тематичної геопросторової інформації на основі картографічних веб-сервісів // Український географічний журнал, 2012. - № 4. – С. 42 – 52, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
55. **Путренко В. В.** Картографування інфраструктурних мереж для цілей управління ризиками виникнення надзвичайних ситуацій // Проблеми безперервної географічної освіти і картографії: Збірник наукових праць. – Харків: ХНУ ім. В.Н. Каразіна, 2012. – Вип. 16. – С. 106 – 109, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
56. **Путренко В. В.** Геоинформационная оценка рисков и опасностей техногенных чрезвычайных ситуаций (на примере Винницкой области) // ИнтерКарто-ИнтерГИС – 18: Устойчивое развитие территорий: теория ГИС и практический опыт. Материалы международной конференции / Редкол.: С.П. Евдокимов (отв. ред.) [и др.]. Смоленск, 26-28 июня, 2012 г. Смоленск, 2012. – С. 367- 374.
57. **Путренко В. В.** Використання хмарних сервісів для тематичного картографування // Ученые записки Таврического национального университета им. В.И. Вернадского, Серия «География» 2012. – Том 25 (64) №1. – С. 191 – 200, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
58. **Путренко В.В.**, Ляшенко Д.О. Математико-картографічне моделювання електоральної ситуації в Україні // Часопис картографії: Збірник наукових

- праць. – К.: КНУ ім. Тараса Шевченка, 2011. – Вип.2. – С. 86-95, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
59. **Путренко В. В.** Підходи до класифікації географічних баз даних // Проблеми безперервної географічної освіти і картографії: Збірник наукових праць. – Харків: ХНУ ім. В.Н. Каразіна, 2011. – Вип.13. – С. 78-81, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
60. **Путренко В. В.** Картографічне моделювання геохімічних полів // Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія: Географічні науки. - 2011 – Т25, №3. – С. 144 – 152, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
61. **Путренко В. В.** Картографічний метод оцінки екологічної безпеки використання підземних вод в Україні // Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія: Географічні науки. - 2011 – Т 24., №2. С.54-62, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
62. **Путренко В.В.,** Голубцов О.Г., Чехній В.М., Фаріон Ю.М. Ландшафтна ГІС як результат ландшафтознавчого прикладного дослідження адміністративного району: методичні аспекти // Географія та туризм, 2010. – Вип. 10. – С. 141 – 153, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
63. **Путренко В.В.,** Півньов О.В Створення та функціонування бази даних Національного атласу України // Укр. геогр. журн. – 2010. – № 1. – С. 53-58, що входить до наукометричних баз Google Scholar, РИНЦ та ін.
64. **Путренко В. В.** Засади створення регіональної ГІС «Ресурсозбереження» // Національне картографування: стан, проблеми та перспективи розвитку: Збірник наукових праць / Відп. За вип. А.А. Москалюк. – К.: ДНВП «Картографія», 2008. – Вип.3 – С. 207 – 212.
65. **Путренко В.В.** Використання глобальних моделей інтерполяції у інтелектуальному аналізі геопросторових даних / В.В. Путренко // Системний аналіз та інформаційні технології: матеріали 17-ї Міжнародної

- науково-технічної конференції SAIT 2015, Київ, 22-25 червня 2015 р. / ННК “ІПСА” НТУУ “КПІ”. – К.: ННК “ІПСА” НТУУ “КПІ”, 2015. – С. 103 – 104.
66. **Путренко В.В.** Інтелектуальний аналіз геопросторових даних для цілей управлінської діяльності / В.В. Путренко // Матеріали Всеукраїнської науково-практичної конференції «В.М. Глушков – піонер кібернетики» (2014 р. м. Київ). – К.: Видавництво «Політехніка», 2014. – С.235 – 237.
67. Putrenko V. Bowkun A., Spatial Modeling of the Risk of Accidents in the Power System of Ukraine // RISK Information Management, Risk Models and Applications, Berlin, Germany, November 17/18, 2014. – P. 10 – 11.
68. **Путренко В.В.** Методи інтелектуального аналізу геопросторових даних // Матеріали XII Міжнародної науково-технічної конференції «Авіа 2015» 28 – 29 квітня 2015 року, –К.: НАУ, 2015. – С. 28.43 – 28.48.
69. **Путренко В.В.,** Джигирей І.М. ГІС-технології як ефективний інструмент залучення громадськості у процедури ухвалення рішень з питань екологічної політики / В.В. Путренко, І.М. Джигирей // Збірка тез доповідей XVII Міжнародної науково-практичної конференції «Екологія. Людина. Суспільство» (м. Київ) / Укладач Д.Е. Бенатов. – К.: НТУУ «КПІ», 2015. – С. 183-185.
70. **Putrenko V.** The Relationship Between Developing OpenStreetMap and Social Activity by the Example of Territory of Ukraine // 27th International Cartographic Conference, Maps Connecting the World, August 23-28, 2015, Rio de Janeiro / Brazil. Processing paper. Web: <http://www.icc2015.org/abstract,517.html>
71. **Putrenko V.,** Govorov M., Gienko G. Unsupervised Machine Learning for Regionalization of Environmental Data: Distribution of Uranium in Groundwater in Ukraine // 27th International Cartographic Conference, Maps Connecting the World, August 23-28, 2015, Rio de Janeiro / Brazil. Processing paper. Web: <http://www.icc2015.org/abstract,892.html>
72. **Путренко В.В.** Статистичний аналіз даних землекористування для цілей регіонального управління в Україні // Міжнародна наукова конференція

Астрономічна школа молодих вчених Україна, Київ, 26–27 травня 2016 р. – К., НАУ. – 2016. – С. 77 – 79.

73. **Putrenko V.** Regional monitoring of Ukraine vegetation changes for yield forecast and climate change monitoring // Copernicus Global Land (Arlon, Belgium, 06-08 June 2016 p.). Web: <http://events.ulg.ac.be/copernicus-gl-workshop/programme/>
74. **Путренко В.В.** Інтелектуальний аналіз геопросторових даних: сучасні тенденції та перспективи // Международный научно-практический форум «НАУКА И БИЗНЕС», 1 июля 2016 года Днепр, Украина. – Днепр, 2016. – С. 76 – 82.
75. **Putrenko V .V., Pashynska N.M.** Wildfire prediction and monitoring in Ukraine on base of Copernicus Land service // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки. – К., С. 41-43
76. **Пашинська Н.М., Путренко В.В.** Використання інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки // Геоінформаційні технології у територіальному управлінні – Одеса : ОРІДУ НАДУ, 2016. – 123-126
77. **Пашинська Н.М., Путренко В.В.** Інтелектуальний аналіз даних на основі класифікації факторів пожежної небезпеки з використанням дерева рішень // Праці VIII міжнародної школи-семінару «Теорія прийняття рішень». – Ужгород, УжНУ, 2016. - С. 203-204
78. **Пашинська Н.М., Путренко В.В.** Використання карт самоорганізації для кластеризації території за рівнем природно-техногенної небезпеки // Міждисциплінарні дослідження актуальних проблем застосування інформаційних технологій в сучасному світі: зб. матеріалів V Всеукр. наук.-практ. конф. "Глушковські читання", Київ, 24 листопада 2016 року. – К: ТОВ НВП "Інтерсервіс", 2016. – С. 145-150.
79. **Viktor Putrenko** Risk Modeling of Accidents in the Power System of Ukraine Based on SDI Data // GSDI 15 Conference Proceedings, November 2016 – p. 134-135 – gsdiassociation.org.

- 80.Пашинська Н.М., **Путренко В.В.** Просторово-часове моделювання забруднення повітря внаслідок аварій на хімічних об'єктах з використанням інструментарію ГІС // Системний аналіз та інформаційні технології: матеріали 19-ї Міжнародної науково-технічної конференції SAIT 2017, Київ, 22 – 25 травня 2017 р. / ННК “ІПСА” НТУУ “КПІ ім. Ігоря Сікорського”. – К.: ННК “ІПСА” НТУУ “КПІ”, 2017. – С. 103-104
- 81.Пархоменко А.В., **Путренко В.В.** Аналіз регіональних збройних конфліктів з використанням методів просторової кластеризації // Системний аналіз та інформаційні технології: матеріали 19-ї Міжнародної науково-технічної конференції SAIT 2017, Київ, 22 – 25 травня 2017 р. / ННК “ІПСА” НТУУ “КПІ ім. Ігоря Сікорського”. – К.: ННК “ІПСА” НТУУ “КПІ”, 2017. – С. 101-102.
- 82.**Путренко В.В.**, Пашинська Н.М. Використання 3D Space-Time Cube для моделювання просторово-часових даних // Збірник матеріалів конференції «ГІС-ФОРУМ-2018» (Харків, 14–16 березня 2018 р.). — Вип. 2. — Х.: ХНУ імені В. Н. Каразіна. – С.32-34.
- 83.Putrenko V., Pashynska N. Clusterization of vegetation dynamics data by method hot spot analysis // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки: Матеріали доповідей. – К., 2018. – С. 42-45
- 84.**Путренко В.В.**, Пашинська Н.М. Оцінка параметрів сталого розвитку методами ДЗЗ // Картографія та вища школа сучасний стан і стратегія розвитку. Матеріали міжнародної конференції. – К., 2018. – С. 71-74
- 85.**Путренко В.В.**, Пашинська Н.М. Використання 3D Space-Time Cube для інтелектуального аналізу даних // System analysis and information technology: 20-th International conference SAIT 2018, Kyiv, Ukraine, May 21- 24, 2018. Proceedings. – pp. 150-151
- 86.Putrenko V., Pashynska N. Analysis of regional armed conflicts using spatial clustering methods // InterCarto-InterGIS 24. Geoinformation and Sustainable Development. Processing - Bonn, July 24-28, 2018. – pp. 31-32

87. **Путренко В.В.** Пашинська Н.М. “Big Data” в просторовому управлінні міським середовищем // Геоінформаційні технології у територіальному управлінні. – Одеса: ОРІДУ, 2015. – С. 89-93
88. **Путренко В.В.** Екологічний моніторинг засобами дистанційного зондування Землі // ЕКОЛОГІЯ ЛЮДИНИ – РОДИНИ – КРАЇНИ: зб. матеріалів круглого столу, (22 листоп. 2018 р., м. Київ) – Київ : КПІ ім. Ігоря Сікорського, 2018. – С. 62 – 64.
89. **Путренко В.В.** Мультимасштабні геопросторові бази даних картографування ризиків та небезпек надзвичайних ситуацій // «Потенціал сучасної географії у розв’язанні проблем розвитку регіонів», Київ 3-5 жовтня 2013 р., с. 432 -435.
90. **Putrenko V.** Multi-scale geospatial database for risk mapping of emergency // From cartography to geographic information science, Vilnius, Litva, May 31 – June
Режим доступу: <http://www.geoportal.lt/wps/poc?uri=page:RUBRIC.2460>.
91. **Путренко В.В.** Засади сервіс-орієнтованого картографування для попередження надзвичайних ситуацій // Матеріали Всеукраїнської конференції «На шляху до сталого розвитку», 18 – 19 червня 2013 р., Київ, с. 120 – 124.
92. **Putrenko V., Rudenko L.** New approaches for mapping of natural, technological and social risks and the dangers of emergency in Ukraine // 32nd International Geographical Congress Book of Abstracts p. 744 Режим доступу <http://igc2012.org/frontend/index.php>
93. **Putrenko V.** GIS-mapping for assessment of risks of groundwater use in Ukraine // 32nd International Geographical Congress Book of Abstracts p. 741 – 742 Режим доступу <http://igc2012.org/frontend/index.php>
94. **Putrenko V.** Risk Management In Ukraine On The Basis Of The Functioning Of Spatial Data Infrastructure // Global Geospatial Conference 2012 - Spatially Enabling Government, Industry and Citizens, Québec City, Canada, 14-17 May 2012, Електронна публікація Режим доступу <http://www.gsdi.org/gsdiconf/gsdi13/abstracts.html#ab82>

95. **Путренко В.В.** Нові технологічні засади екологічного картографування // Зелена економіка: перспективи впровадження в Україні: матеріали міжнародної конференції (Київ, 24 – 25 квітня 2012 р.): у 3 т. – К. : Центр екологічної освіти та інформації, 2012. – Т.1. – С. 369 – 372.
96. **Путренко В.В.,** Чабанюк В.С., Дишлик О.П. Головні технологічні рішення при створенні національних атласів // Національне картографування: стан, проблеми та перспективи розвитку, Збірник матеріалів V Всеукраїнської науково-практичної конференції «Національні атласи у формуванні глобального інформаційного простору» (м. Київ, 13-14 вересня 2012 р.), Вип. 5. – 2012. – С. 177 – 180
97. **Путренко В.В.,** Руденко Л.Г. Атлас ризиків та небезпек виникнення надзвичайних ситуацій // Національне картографування: стан, проблеми та перспективи розвитку, Збірник матеріалів V Всеукраїнської науково-практичної конференції «Національні атласи у формуванні глобального інформаційного простору» (м. Київ, 13-14 вересня 2012 р.), Вип. 5. – 2012. – С. 155 – 159.
98. **Путренко В.В.,** Руденко Л.Г., Волошин С.М. Інформаційний продукт щодо аналізу ризиків та небезпек виникнення надзвичайних ситуацій в Україні // Екологічні аспекти регіонального партнерства в надзвичайних ситуаціях: Збірка матеріалів I Міжвузівської науково-методичної конференції. – Х.: НУЦЗУ, 2012. – С. 97 – 103.
99. **Путренко В.В.** Геоінформаційний аналіз регіональної інфраструктури з метою попередження надзвичайних ситуацій // Геопросторові технології в сталому розвитку міст: матеріали Міжвузівської науково-практичної конференції, Харків, - 16- 17 листопада 2011 р. – Х.: ХНАМГ, 2011. – С. 16-18
100. **Putrenko V.** Cartographic modeling of differentiation electoral space in Ukraine // Territorial identity and geopolitics (18-23 september 2011, Kyiv-Crimea, Ukraine). – Kyiv, 2011. – P. 33 – 36

101. **Путренко В.В.** Світовий досвід організації тематичної інформації у інфраструктурах геопросторових даних // Розвиток тематичної складової інфраструктури геопросторових даних в Україні: Збірник наукових праць. – К., 2011. – С. 133 – 139
102. **Путренко В.В.** Побудова геоінформаційних баз даних управління природними ресурсами // Природно-ресурсний потенціал збалансованого (сталого) розвитку України: Матеріали Міжнародної науково-практичної конференції (Київ, 19-20 квітня 2011 р.) у 2 т. – К.: Центр екологічної освіти та інформації, 2011. – Т.1. – С. 424 – 427
103. **Путренко В.В.** Базы данных тематического картографирования в составе инфраструктур пространственных данных // Тематическое картографирование для создания инфраструктур пространственных данных / Материалы IX научной конференции по тематической картографии (Иркутск, 9-12 ноября 2010 г.). – Иркутск: Изд-во Института географии им. В.Б. Сочавы СО РАН, 2010. – В 2-х т. – Т. 1. –С. 56 – 59.
104. **Путренко В.В.** База даних ландшафтознавчого дослідження адміністративного району з метою моніторингу та раціонального використання земель
105. **Путренко В.В.** Підходи до формування бази даних атласу надзвичайних ситуацій // Просторовий аналіз природних і техногенних ризиків в Україні: Зб. Наук. праць. – 2009. – С. 37 – 40.
106. **Путренко В.В.** Геоінформаційне атласне картографування природокористування // Українська історична географія та історія географії в Україні: Матеріали міжнародної наукової конференції (7-10 жовтня). Чернівці: Чернівецький національний університет, – 2009. – С. 176.
107. **Putrenko V.V., Pashynska N.M.** Web Atlas of Sustainable Development of Ukraine // ICA Commissions Joint Workshop on Atlases, Cognition, Usability. Olomouc, April 27–30, 2018. – p. 28-29.

108. **Putrenko V.**, Modeling Spatial Relationships of Electoral Behavior in Ukraine
// 25th World Congress of Political Science, July 21-25, 2018, Brisbane, Australia
<https://www.ipsa.org/events/congress/brisbane2018>

ЗМІСТ

	стор.
ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ	38
ВСТУП	40
РОЗДІЛ 1. СИСТЕМНИЙ ПІДХІД ДО АНАЛІЗУ ГЕОПРОСТОРОВИХ ДАНИХ	51
1.1. Розвиток концепції та метрика сталого розвитку	51
1.2. Геопросторові дані як об'єкт аналізу	61
1.3. Джерела, методи збору та інтеграції геопросторових даних	73
1.4. Системний підхід та методологія обробки геопросторових даних для задач сталого розвитку	83
Висновки до розділу	97
РОЗДІЛ 2. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ НА ОСНОВІ ГЕОПРОСТОРОВИХ ДАНИХ	100
2.1. Базові моделі представлення геоданих	100
2.2. Базові методи інтелектуального аналізу геопросторових даних	108
2.3. Методи просторового аналізу гарячих точок	118
2.4. Моделювання просторових відносин	125
2.5. Метод просторово-часового кубу для аналізу геопросторових даних	131
Висновки до розділу	140
РОЗДІЛ 3. МЕТОДИ КЛАСТЕРИЗАЦІЇ ГЕОПРОСТОРОВИХ ДАНИХ	142
3.1. Огляд класичних методів класифікації багатовимірних даних	142
3.1.1 Постановка завдання класифікації багатовимірних даних	142
3.1.2 Модель суміші розподілів	144
3.1.3 Методи класифікації без навчання	146
3.2. Класифікації багатовимірних даних	152
3.2.1 Оціночні і типологічні класифікації багатовимірних даних	152
3.2.2 Алгоритми класифікації багатовимірних даних, що застосовуються в геоінформації	153

3.2.3 Облік при класифікації багатовимірних даних географічного простору	156
3.3 Реалізація методів автоматичної класифікації і районування в сучасних векторних ГІС	161
3.3.1 Реалізація в векторних ГІС методів класифікації за єдиною ознакою	161
3.3.2 Реалізація в векторних ГІС методів класифікації за багатьма ознаками	162
3.4 Ядерні методи автоматичного районування	164
3.4.1 Загальна схема ядерного алгоритму районування	164
3.4.2 Алгоритм районування на основі методу класифікації k-методів	165
3.4.3 Ізоморфізм постановок задач класифікації багатовимірних даних при наявності обмежень	166
3.5. Ієрархічні методи автоматичного районування	167
3.5.1 Агломеративні ієрархічні алгоритми районування	167
3.5.2 Дивізімні ієрархічні алгоритми районування на основі класичних дивізімних методів класифікації	169
3.5.3 Дивізімний ієрархічний алгоритм районування бар'єрів максимальних відмінностей	170
3.6. Отримання, зберігання та використання матриці суміжності	173
3.7. Методи багатовимірної кластеризації даних	176
3.8. Алгоритм нечіткої самоорганізації k-середніх	185
3.9. Розробка алгоритму побудови моделі сукупності однорідних територіальних зон	187
3.10. Кластеризація та регіоналізація геопросторових даних в управлінні	189
3.10.1. Задачі управління екологічними небезпеками	189

3.10.2. Регіоналізація території України за результатами кластеризації у просторі показників сталого розвитку	197
3.11. Геопросторовий аналіз даних з використанням методу гарячих точок	202
3.11.1. Кластеризація геопросторових даних про військову конфліктогенність у світі	202
3.11.2. Аналіз динаміки вегетаційного стану рослинності	207
Висновки до розділу	212
РОЗДІЛ 4 ГЕОСТАТИСТИЧНЕ МОДЕЛЮВАННЯ В АНАЛІЗІ ДАНИХ	216
4.1. Проблеми просторового моделювання	216
4.2. Основні поняття та елементи геостатистики	226
4.3. Детерміновані методи просторової інтерполяції	239
4.4. Геостатистичні інтерполяції для скалярної функції методом крігінгу	246
4.5. Застосування трендових геопросторових статистичних моделей для оцінки сталого регіонального розвитку	258
4.6. Аналіз якості методів інтерполяції даних з використанням мови R	269
Висновки до розділу	294
РОЗДІЛ 5. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ГЕОПРОСТОРОВИХ ДАНИХ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ	296
5.1. Технології штучного інтелекту в ГІС	296
5.2. Застосування нейронних мереж без вчителя для аналізу геопросторових даних	305
5.3. Параметризація складових сталого розвитку з використанням показників нічної освітленості в Україні	317
5.4. Застосування нейронної мережі для завдань регіонального аналізу індексу економічного розвитку в моделі сталого розвитку	325
5.5. Реалізація елементів СППР інтелектуального аналізу геопросторових даних для задач сталого розвитку	343

Висновки до розділу	353
ВИСНОВКИ	357
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	362
Додаток А Індикатори моделювання сталого розвитку регіонів України	403
Додаток Б Діагностика інтерполяційних геостатистичних моделей розподілу значень миш'яку у підземних водах території України	435
Додаток В Копії актів впровадження	444

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ГДК – гранично допустима концентрація

ГПІ – геоінформаційний простір

ГІС – географічна інформаційна система

ГМА – географічна машина аналізу

ДЗЗ – дистанційне зондування Землі

ІАГД – інтелектуальний аналіз геопросторових даних

ІПД – інфраструктура просторових даних

ЛДА – лінійний дискримінантний аналіз

ЛІПД – локальна інфраструктура просторових даних

ММП – метод максимальної правдоподібності

ОГІ – образно-географічних інтерпретацій

ООН – Організація об’єднаних націй

ОТО – окрема територіальна одиниця

ППХ – повна просторова хаотичність

РЛСБО – радіолокаційна система бічного огляду

СКБД – система управління базою

СППР – система підтримки прийняття рішень

ЦКЗ – цифрових картографічних зображень

ЦРТ – Цілі Розвитку Тисячоліття

ІНМ – штучна нейронна мережа

API – Application Programming Interface

EDA – дослідницький аналіз даних

ESDA – пошуковий просторовий аналіз даних

GNSS – Global navigation satellite system

GPS – Global Positioning System

IP – Internet Protocol address

LAI – Leaf Area Index

LIDAR – Light Identification, Detection and Ranging

NAVSTAR – Navigation Satellite Time and Ranging

NDVI – нормалізований відносний індекс рослинності

NIR – відображення в ближній інфрачервоній області спектра

NSDI - інфраструктура просторових даних

PCA – аналіз головних компонент

RED - відображення в червоній області спектра

RFID – радіочастотна ідентифікація

RSS – Really Simple Syndication

VCI – Vegetation Condition Index

VPI – Vegetation Productivity Index

W3C – World Wide Web Consortium

ВСТУП

Актуальність теми. Підтримка прийняття управлінських рішень в територіально розподілених системах засновується на використанні геопросторової інформації, що в свою чергу вимагає застосування складних математичних та обчислювальних інструментів і методів інтелектуального аналізу та обробки геопросторових даних з метою досягнення цілей сталого розвитку та безпеки життя людей на відповідних територіях.

Зазначені тенденції дають поштовх для розвитку індустрії геопросторових даних, переходу до управління територіями (велике місто, регіон країни, країна) на основі використання геоінформаційних систем, створення нових систем та методів збору геопросторових даних, в тому числі геокодування та геотегування інформації, накопичення великих обсягів геопросторових даних. Ці задачі вимагають застосування новітніх способів аналізу великих обсягів геопросторових даних з метою виявлення прихованих закономірностей в їх структурі. Тому інтелектуальний аналіз геопросторових даних, як важливий напрям наукових досліджень, відкриває нові можливості для побудови ефективних систем територіального управління.

Попередні дослідження з цієї тематики викладені в роботах: А.Г. Лексин, В.В. Швецов, Н.В. Попович, А.Д. Гвішиані, М.З. Згуровський, О.І. Шаблій, James Johannesson, С.Н. Березко, Л.Б. Рыбкина, Diansheng Guo, Jeremy Mennis, Andrienko & Andrienko, 1999; Chawla et al., 2000; Gahegan, 2003; Guo, Peuquet, & Gahegan, 2003; Guo et al., 2006; Han, Koperski & Stefanovic, 1997; Keim et al., 2004; Knorr & Ng, 1996; Kulldorff, 1997; Mennis & Liu, 2005; Miller & Han, 2009; Miller and Han, 2001; Openshaw, Charlton, Wymer, & Craft, 1987; Shekhar et al. 2004.

Разом з тим, існуючі на сьогодні методи інтелектуального аналізу великих даних мають суттєві обмеження в частині їх практичного застосування до обробки геопросторових даних. З іншого боку, кількісні методи аналізу геопросторових даних не можуть надати інтегральної оцінки для формування регіональної політики

стосовно досягнення цілей сталого розвитку без розробки відповідної системної методології сталого розвитку з безпековою складовою.

Таким чином, розробка якраз такої системної методології та відповідних математичних і обчислювальних інструментів інтелектуального аналізу геопросторових даних для задач сталого розвитку великих територіальних об'єднань зумовлює актуальність проблеми дисертаційної роботи.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана відповідно до плану наукових досліджень у рамках таких науково-дослідних програм (тем) ННК «Світовий центр даних з геоінформатики та сталого розвитку» та ННК «ІПСА» КІІ ім. Ігоря Сікорського:

- Розробка методології та математичної бази кількісної оцінки процесів сталого розвитку та впливу на них сукупності загроз як у глобальному, так і в регіональному контексті (0111U001738)
- Форсайт економіки України: середньостроковий (2015–2020 роки) і довгостроковий (2020–2030 роки) часові горизонти
- Міжнародний проект EU funded Programme for the prevention, preparedness, and response to man-made and natural disasters in the ENPI East Region – EuropeAid/129397/C/SER/Multi (EP 105044)
- Міжнародний проект DFATD funded project «Laying the Foundation for a Spatial Data Infrastructure: Building Capacity within the Ukrainian Government to Support Sustainable Economic Growth» (S-065681)
- Аналіз соціально-економічних процесів розвитку суспільства, заснованого на знаннях (0114U000670)
- Розробка он-лайн платформи аналізу і сценарного планування сталого розвитку регіонів України в контексті якості та безпеки життя людей (0117U002476)
- Міжнародний проект 6166 спільно з УНТЦ та НАН України «Розробка інформаційних технологій для моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів та розповсюдження

зброї на стабільність розвитку суспільства в регіональному та глобальному масштабах» (STCU #6166)

Мета роботи: Розробка методології та сукупності математичних і обчислювальних інструментів інтелектуального аналізу геопросторових даних в системах підтримки прийняття управлінських рішень для великих територіальних об'єднань з метою сталого розвитку та підвищення якості і безпеки життя людей на зазначених територіях.

Завдання:

1. Проаналізувати існуючі методи та підходи до моделювання, оцінювання і прогнозування складових сталого розвитку на основі геопросторових даних.
2. Системно обґрунтувати особливості методів інтелектуального аналізу геопросторових даних в задачах сталого розвитку в межах концепції побудови вектору сталого розвитку на основі кватерного підходу до побудови компонентів якості життя та глобальних загроз.
3. Розробити методологію інтелектуального аналізу геопросторових даних в системах підтримки прийняття рішень для задач сталого розвитку з обґрунтуванням вибору геоінформаційної платформи для інтелектуального аналізу геопросторових даних.
4. Дослідити застосування геоінформаційних технологій для проведення інтелектуального аналізу з використанням основних моделей організації геопросторових даних.
5. Розробити систему методів, моделей та алгоритмів інтелектуального аналізу для обробки геопросторових даних для вирішення задач визначення впливу геопросторових відносин на складові сталого розвитку методом побудови трендових поверхонь.
6. Удосконалити методи стохастичного моделювання невідомих значень у геопросторі шляхом модифікації варіаграмної моделі крігінгу для відновлення відсутніх значень у статистичних рядах метрики сталого розвитку.
7. Запропонувати методи інтелектуального аналізу геопросторових даних на основі ГІС-технологій для аналізу безпекового компоненту на різних

територіальних рівнях методами побудови просторово-часових кубів спільно з аналізом кластеризації «гарячих точок».

8. Розробити методи кластеризації регіонів України у багатовимірному просторі з врахуванням просторових відносин для проведення районування та зонування території.
9. Удосконалити методи глобальної та локальної просторової регресії для вирішення завдань оптимізації конфігурування просторових патернів у розподілі точкових об'єктів та методи методи геопросторової концептуалізації топологічних відносин між об'єктами
10. Розробити методи розрахунку метрики сталого розвитку на основі даних дистанційного зондування Землі з використанням методів карт самоорганізації Кохонена та згорткових нейронних мереж з градієнтним алгоритмом навчання.
11. Розробити інструментарій СППР на основі інтелектуального аналізу геопросторових даних та нейронних мереж з метою впровадження науково-обґрунтованого підходу до управління на регіональному та локальному рівні.
12. Застосувати розроблені інформаційні технології та системи для розв'язання прикладних задач моделювання параметрів сталого розвитку регіонів України.

Об'єкт дослідження: Процеси дослідження інтелектуального аналізу даних (Data Mining) з геопросторовим положенням.

Предмет дослідження: Моделі, методи та технології інтелектуального аналізу геопросторових даних для використання в системах прийняття рішень в задачах сталого розвитку великих територіальних об'єднань.

Методи дослідження ґрунтуються на використанні методології системного аналізу, методів і алгоритмів передбачення, теорії підтримки прийняття рішень та геоінформаційного підходу, інтелектуальних методів аналізу геопросторових даних, зокрема, методів класифікації та кластеризації геопросторових даних, асоціативних правил, зонування та районування, геостатистики та географічної регресії, побудови просторово-часових кубів, методах штучного інтелекту, в тому числі багатошарових нейронних мережах.

Наукова новизна: У дисертації сформульовано та обґрунтовано низку наукових положень, які у своїй сукупності дозволяють розв'язати важливу науково-прикладну проблему інтелектуального аналізу геопросторових даних з метою розпізнавання прихованих закономірностей та відношень в структурі цих даних при вирішенні задач сталого розвитку великих територіально розподілених систем.

Основні наукові результати дисертаційної роботи полягають у наступному:
вперше:

- розроблено методологію використання інтелектуальних методів аналізу геопросторових даних, яка включає в себе принципи сталого розвитку, моделі, методи та технології системного управління територіями;
- розроблено методику використання геопросторових даних дистанційного зондування Землі для отримання інтегральної оцінки параметрів сталого розвитку та безпеки життя людей:
 - розроблено метод моделювання на основі географічно зваженої регресії та локального індексу Морана;
 - розроблено метод розрахунку індексу економічного розвитку території з використанням нейронної мережі на основі навчання на базі даних дистанційного зондування Землі.
- розроблено методику трендового аналізу функціонального та кількісного зв'язку вкладу складових індексу сталого розвитку
- розроблено методику районування території за параметрами сталого розвитку з використанням топологічних геопросторових даних місцеположення з використанням методів геопросторової класифікації та кластеризації

удосконалено:

- компонент безпеки моделі сталого розвитку на основі індексу природних та техногенних небезпек.
- методи геостатистичного аналізу та оптимізації задачі інтерполяції даних стохастичними методами крігінгу.

- методи аналізу просторово-часого кубу даних у дослідженні великих геопросторових даних.

отримали подальший розвиток:

- методи просторової кластеризації та нечіткої кластеризації даних з використанням статистичної обробки «гарячих-холодних» точок
- методи геопросторової концептуалізації топологічних відносин між об'єктами
- методи глобальної та локальної просторової регресії для вирішення завдань оптимізації конфігурування просторових паттернів у розподілі точкових об'єктів.

Практичне значення одержаних результатів полягає у створенні інструментарію на основі програмної платформи ArcGIS для інтеграції методів інтелектуального аналізу та нейронних мереж до геоінформаційної системи з метою оцінки параметрів сталого розвитку території з позицій кластеризації, районування та комплексування адміністративно-територіальних одиниць для досягнення цілей сталого розвитку та безпеки життя людей. Розроблений підхід із застосування нейронних мереж дозволяє оцінити параметри та інтегральну оцінку складових сталого розвитку на різних територіальних рівнях, які відповідають європейській класифікації регіонів NUTS 1-5 та державного адміністративно-територіального устрою України, включаючи новий тип адміністративних територіальних утворень – об'єднані територіальні громади.

Ще одним аспектом застосування інструментарію є можливість отримання необхідних даних для територій, на яких статистичні дані не збираються. В Україні до класу таких територій відноситься окупована частина Донбасу та АР Крим.

Побудований інструментарій використано при розв'язанні наступних практичних задач:

- оцінка просторових трендів розподілу регіонів України за індикаторами сталого розвитку за період з 2013 р. по сьогоднішній день;
- побудова схем районування території України в залежності від особливостей розподілу індикаторів сталого розвитку;

- класифікація адміністративних одиниць України за структурою землекористування.

- оцінювання індексу економічного розвитку як складової якості життя населення для територіальних одиниць України та його динаміки з 2013 р.

Результати дисертаційної роботи впроваджено в проектах ННК «Світовий центр даних з геоінформатики та сталого розвитку» в межах дослідницьких робіт та ініціативних тематик аналізу параметрів сталого розвитку країн світу та України. Крім того, результати дисертації були впроваджені в Інституті телекомунікації та глобального простору Національної академії наук України, Державному підприємстві "Український державний науково-дослідний інститут проектування міст "ДІПРОМІСТО" імені Ю.М.Білоконя.

Результати дисертаційної роботи впроваджено в навчальний процес кафедри «Математичні методи системного аналізу» Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», зокрема при викладанні дисципліни «Геоінформаційні системи» та «Інтелектуальний аналіз в геоінформаційних системах».

Особистий внесок автора у праці, які склали основу дисертації і виконані в співавторстві.

Всі наукові положення та результати, що складають зміст роботи та становлять наукову новизну, отримані автором самостійно. Робота містить теоретичні та методичні положення, висновки, які сформульовані дисертантом особисто.

Здобувачем поставлено і розв'язано задачі розробки методології інтелектуального аналізу геопросторових даних, застосування методів інтелектуального аналізу геопросторових даних для аналізу параметрів сталого розвитку регіонів України на основі методології використання геопросторових даних дистанційного зондування Землі для отримання синтетичної оцінки параметрів сталого розвитку та безпеки життя людей, методу моделювання на

основі географічно зваженої регресії, глобального та локального індексу Морана, методу розрахунку індексу економічного розвитку території з використанням багатословних нейронних мереж на основі навчання на базі даних дистанційного зондування Землі. Здобувачем розроблено підходи на базі геостатистичного аналізу даних до комплексного оцінювання індикаторів сталого розвитку на прикладі регіонів України.

Здобувачем запропоновано ідею та методологію отримання оціночних показників сталого розвитку території з використанням знімків дистанційного зондування Землі в якості вхідних даних для навчання нейронної мережі, яка реалізована за допомогою бібліотеки на мові Python у середовищі ArcGIS 10.4.

Апробація результатів дисертації.

- міжнародна наукова конференція «Українська історична географія та історія географії в Україні» (м. Чернівці, 7-10 жовтня).
- міжнародна наукова конференція «Просторовий аналіз природних і техногенних ризиків в Україні» (м. Київ, 2009)
- міжнародна наукова конференція «IX конференция по тематической картографии» (Иркутск, 2010);
- міжнародна науково-практична конференція «Природно-ресурсний потенціал збалансованого (сталого) розвитку України» (м. Київ, 2011)
- міжнародний науковий семінар «Розвиток тематичної складової інфраструктури геопросторових даних в Україні» (м. Київ, 2011)
- міжнародна наукова конференція «Territorial identity and geopolitics» (м. Київ, 2011)
- міжвузівська науково-практична конференція «Геопросторові технології в сталому розвитку міст» (Харків, 2011)
- міжвузівська науково-практична конференція «Екологічні аспекти регіонального партнерства в надзвичайних ситуаціях» (Харків, 2012)
- V Всеукраїнська науково-практична конференція «Національні атласи у формуванні глобального інформаційного простору» (Київ, 2012)

- міжнародна наукова конференція «Зелена економіка: перспективи впровадження в Україні», (Київ, 2012)
- міжнародна наукова конференція «Global Geospatial Conference 2012 - Spatially Enabling Government, Industry and Citizens» (м. Квебек, Канада, 2012)
- міжнародна наукова конференція «32nd International Geographical Congress» (Кельн, Німеччина, 2012)
- міжнародна наукова конференція «На шляху до сталого розвитку» (Київ, 2013)
- міжнародна наукова конференція «From cartography to geographic information science» (Вільнюс, Литва, 2013)
- міжнародна наукова конференція «Потенціал сучасної географії у розв'язанні проблем розвитку регіонів» (Київ, 2013)
- міжнародна наукова конференція «Геоінформаційні технології у територіальному управлінні» (Одеса, 2015, 2016)
- міжнародна наукова конференція «Астрономічна школа молодих вчених України» (Київ, 2016 р.)
- міжнародна наукова конференція «InterCarto-InterGIS 24. Geoinformation and Sustainable Development» (Бонн, Німеччина, 2018)
- міжнародна наукова конференція «International conference SAIT» (Київ, 2014, 2015, 2016, 2017, 2018)
- міжнародна наукова конференція «Картографія та вища школа сучасний стан і стратегія розвитку» (Київ, 2018)
- міжнародна наукова конференція «Аерокосмічні спостереження в інтересах сталого розвитку та безпеки» (Київ, 2016, 2018)
- міжнародна наукова конференція «ГІС-ФОРУМ-2018» (Харків, 2018).
- міжнародна наукова конференція «GSDI 15 Conference» (Тайбей, Тайвань, 2016)
- всеукраїнська науково-практична конференція «Глушковські читання» (Київ, 2016)

- VIII міжнародна школа-семінар «Теорія прийняття рішень» (Ужгород, 2016)
- Міжнародний науково-практичний форум «НАУКА И БИЗНЕС» (Дніпро, 2016)
- міжнародна наукова конференція «Copernicus Global Land» (Арлон, Бельгія)
- міжнародна наукова конференція «27th International Cartographic Conference» (Ріо-де-Жанейро, Бразилія)
- міжнародна наукова конференція «Екологія. Людина. Суспільство» (Київ, 2015)
- міжнародна наукова конференція «Авіа 2015» (Київ, 2015)
- міжнародна наукова конференція «RISK Information Management, Risk Models and Applications» (Берлін, Німеччина)
- всеукраїнська науково-практична конференція «В.М. Глушков – піонер кібернетики» (Київ, 2014)
- міжнародний науковий семінар ICA Commissions Joint Workshop on Atlases, Cognition, Usability (Оломоуц, Чеська республіка, 2018)
- міжнародна наукова конференція «25th World Congress of Political Science» (Брісбен, Австралія, 2018)
- всеукраїнська науково-практична конференція «Моделювання та прогнозування економічних процесів» (Київ, 2017)
- всеукраїнська науково-практична конференція «Сучасні проблеми управління: імперативи забезпечення сталого розвитку» (Київ, 2017)
- міжнародна наукова конференція «The 28th International Cartographic Conference: ICC 2017» (Вашингтон, США, 2017)
- міжнародна наукова конференція «IEEE First Ukraine Conference on ELECTRICAL AND COMPUTER ENGINEERING (UKRCON)» (Київ, 2017)

Публікації. За темою дисертації опубліковано 108 праці, у тому числі 11 монографій, 24 статей у фахових виданнях (8 – у виданнях, включених до наукометричної бази даних Scopus, 5 – у міжнародних виданнях, 1 авторське

свідомство), 63 – у матеріалах наукових конференцій та інших виданнях, 3 навчальні посібники.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, п'яти розділів, висновків, списку використаних джерел із 416 найменувань та трьох додатків на 40 сторінках. Загальний обсяг дисертації становить 449 сторінок, у тому числі 327 сторінок основного тексту, робота містить 115 рисунків та 16 таблиць.

РОЗДІЛ 1. СИСТЕМНИЙ ПІДХІД ДО АНАЛІЗУ ГЕОПРОСТОРОВИХ ДАНИХ

1.1 Розвиток концепції та метрика сталого розвитку

Доповідь «Наше спільне майбутнє», яка була підготовлена під керівництвом Г.Х. Брундланд міжнародною комісією з навколишнього середовища та розвитку, міжнародна конференція ООН в Ріо-де-Жанейро в 1992 році, всесвітній самміт в 2002 р. в Йоганнесбурзі стали основними поворотними віхами в становленні та розвитку концепції сталого розвитку. Причиною створення концепції сталого розвитку є необхідність подолання глобальної кризи, яка назріла у відносинах між суспільством та природою і загрожує продовженню розвитку людства та біосфери Землі. Глобальний характер проблем, з якими зіткнулось людство, вимагає розробки і впровадження наддержавної програми дій, в реалізації якої повинні взяти участь всі країни світу. Концепція сталого розвитку може бути ефективною, коли стане засадничою доктриною державного управління в світі та буде враховувати регіональні та локальні особливості розвитку країн світу.

Головним завданням концепції сталого розвитку є збереження природного середовища та ресурсів для майбутніх поколінь, а це передбачає зменшення споживання природних ресурсів, збереження та захист від надмірного антропогенного навантаження екосистеми Землі.

Основними напрямками реалізації концепції стали екологічний, економічний та соціальний. Тільки досягнення збалансованості розвитку по усіх цих параметрах може призвести до стабільності загального розвитку людства. Ключовою ланкою стає взаємовідношення між людством та навколишнім середовищем. Тому необхідне впровадження програм переходу до раціонального споживання, відходу від неконтрольованої експлуатації природних ресурсів, загальне зменшення кількості забруднюючих речовин, що потрапляють у навколишнє середовище. Це вимагає трансформації економіки країн, що розвиваються, переходу від переважання галузей добувної промисловості до збільшення частки виробництв з високим ступенем переробки, що не можливо без науково-технічного співробітництва з розвинутими країнами. В іншому випадку, це призведе до

економічної нестабільності, а за тим, неминуче – до розгляду економічного аспекту сталого розвитку, який полягає у вирівнюванні різниці в економічному розвитку між державами світу, та загального переходу світової економіки до постіндустріального етапу розвитку з переважанням наукоємних виробництв, сфери послуг, замкнених циклів виробництва, що попереджають забруднення.

З покращенням екологічної та економічної ситуації безпосередньо пов'язана соціальна складова збалансованого розвитку. Вирішення екологічних проблем, забезпечення населення якісним водопостачанням та продуктами харчування, підвищення добробуту людей, забезпечення медичним обслуговуванням, організація освіти є негачними завданнями збалансованого розвитку, реалізація яких не можлива без кардинальних змін екологічної та економічної ситуації.

Серйозними перешкодами на шляху до сталого розвитку є політична нестабільність у світі, значна різниця у рівні економічного розвитку між країнами, яка за останнє десятиліття лише збільшилась, диспропорція у розподілі природних, трудових, фінансових та інформаційних ресурсів, наявність протиставлення між високим рівнем споживання в розвинутих країнах, зменшення якого є проблематичним, та нестачею ресурсів для підтримання біологічних потреб населення в країнах, що розвиваються [148].

До основних механізмів впровадження концепції сталого розвитку слід віднести визначення показників збалансованості розвитку на основі розробки відповідних індексів, вивчення та удосконалення регіональних механізмів природокористування на основі досвіду місцевих етносів, розвиток системи екологічної освіти, розбудову механізмів міждержавного управління.

«Декларація тисячоліття» ООН, прийнята у 2000 році 189-ма країнами на Саміті тисячоліття ООН, визначила Цілі Розвитку Тисячоліття (ЦРТ), тобто всеосяжні рамки цінностей, принципів і ключових чинників розвитку до 2015 року. Після розроблення універсальних рамок ЦРТ у багатьох країнах світу було здійснено адаптацію ЦРТ та встановлено цільові показники розвитку з урахуванням специфіки національної ситуації. Україна приєдналася до «Декларації тисячоліття» ООН [250]:

Ціль 1. Подолання бідності.

Ціль 2. Подолання голоду, розвиток сільського господарства.

Ціль 3. Міцне здоров'я і благополуччя.

Ціль 4. Якісна освіта.

Ціль 5. Гендерна рівність.

Ціль 6. Чиста вода та належні санітарні умови.

Ціль 7. Доступна та чиста енергія.

Ціль 8. Гідна праця та економічне зростання.

Ціль 9. Промисловість, інновації та інфраструктура.

Ціль 10. Скорочення нерівності.

Ціль 11. Сталий розвиток міст і громад.

Ціль 12. Відповідальне споживання та виробництво.

Ціль 13. Пом'якшення наслідків зміни клімату.

Ціль 14. Збереження морських ресурсів.

Ціль 15. Захист та відновлення екосистем суші.

Ціль 16. Мир, справедливість та сильні інститути.

Ціль 17. Партнерство заради сталого розвитку [258].

Ста́лий ро́звиток (англ. Sustainable development) — загальна концепція стосовно необхідності встановлення балансу між задовільненням сучасних потреб людства і захистом інтересів майбутніх поколінь, включаючи їх потребу в безпечному і здоровому довкіллі. Сталий розвиток — це систематично керований розвиток. Основою його керованості є системний підхід та сучасні інформаційні технології, які дозволяють дуже швидко моделювати різні варіанти напрямків розвитку, з високою точністю прогнозувати їх результати та вибрати найбільш оптимальний. Концепція сталого розвитку з'явилася в результаті об'єднання трьох основних точок зору: економічної, соціальної та екологічної [258].

Принципи сталого розвитку базуються на розумінні розвитку як процесу, при якому нинішні покоління задовольняють свої потреби, при цьому, не ставлячи під загрозу, можливість задовольняти свої потреби майбутнім поколінням.

До числа основних принципів сталого розвитку також відноситься:

- принцип системності та комплексності розгляду розвитку в системі людина – природа;
- принцип гармонійності розвитку соціальної, економічної та екологічної складових сталого розвитку;
- принцип балансу споживання природних ресурсів за для запобігання їх вичерпанню.

Введено новий принцип оперування геопросторовими даними, за яким усі системні зв'язки розглядаються у їх геопросторових залежностях на певній території. На основі цього принципу, методології системного аналізу та методології оцінювання сталого розвитку [5, 382] розроблено методологію інтелектуального аналізу геопросторових даних для задач сталого розвитку.

Наведена методологія ґрунтується на формалізації процесів розвитку складних соціально-економічних систем в тривимірному географічному просторі, який впливає на зміну параметрів процесів за рахунок тісноти просторових зв'язків, топологічних відносин та взаємодії об'єктів. Метрика взаємодії у геопросторі задається через координатні описи об'єктів в певній системі геопросторових координат та досліджується з використанням сучасних методів інтелектуального аналізу даних та розроблених методів та методик аналізу геопросторових даних в задачах сталого розвитку.

Для обробки великих обсягів геопросторових даних в дисертації використовується SDGM – метод (Sustainable Development Gauging Matrix), запропонований в роботах Згуровського М.З. [414 – 415].

Відповідно до SDGM метода вводиться функціональна залежність:

$$G_{sd} = \langle \{Q\}(x, y, z, t) \rangle,$$

де G_{sd} – геопросторова складова вимірювання метрики сталого розвитку; $\{Q\}$ – кватерніон міри сталого розвитку, x – координата широти, y – координата довготи у визначеній геодезичній системі координат, z – координата значення в точці в умовній системі координат, t – часова змінна у розвитку процесу сталого розвитку.

Таким чином, згідно SDGM метода міра сталого розвитку розглядається в тривимірному просторі в динамічній зміні, що впливає на її причинно-наслідкові

зв'язки та прогнозування цих змін, а сам процес сталого розвитку характеризується двома основними складовими: безпекою (C_{sl}) і якістю (C_{ql}) життя людей. Узагальнена міра сталого розвитку визначається за допомогою кватеріона [414 – 415] (рис. 1.1):

$$\{Q\} = j w_{sl} C_{sl} + w_{ql} \vec{C}_{ql}(I_{ec}, I_e, I_s). \quad (1.1)$$

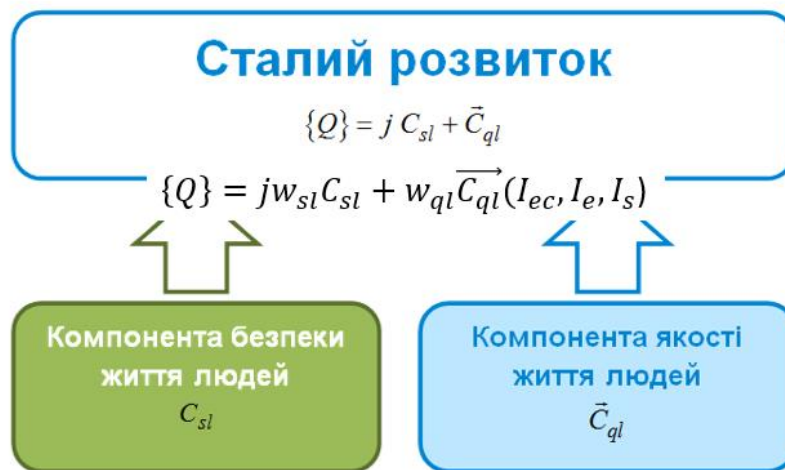


Рис. 1.1. Моделювання процесу сталого розвитку на основі компонентів безпеки та якості життя населення.

Кватеріон $\{Q\}$ містить уявну зважену скалярну частину $j w_{sl} C_{sl}$, яка описує безпеку життя людей і зважену дійсну векторну частину, яка описує якість життя людей у просторі трьох вимірів: економічного (I_e), екологічного (I_{ec}) і соціально-інституціонального (I_s). При цьому j набуває значення дійсної одиниці для нормального, регулярного стану розвитку суспільства за $C_{sl} > 0$, і значення уявної одиниці, коли суспільство переходить у стан конфлікту ($C_{sl} = 0$) [415]:

$$j = \begin{cases} 1, & \text{для } C_{sl} > 0; \\ \sqrt{-1}, & \text{для } C_{sl} = 0 \text{ (конфлікт)}. \end{cases}$$

Вагові коефіцієнти w_{sl} і w_{ql} в формулі (1.1) використані з метою вирівнювання масштабів компонент безпеки і якості життя (у випадку оцінювання сталого розвитку регіонів України – $w_{sl} = w_{ql} = 1$).

Таким чином, відповідно до SDGM – метода, під *індексом сталого розвитку* будемо розуміти кількісну міру сталого розвитку, яка враховує безпеку та якість життя людей та для випадку $C_{sl} > 0$ обчислюється як норма кватеріону $\{Q\}$:

$$\|\{Q\}\| = \sqrt{w_{sl}^2 C_{sl}^2 + w_{ql}^2 (I_e^2 + I_{ec}^2 + I_s^2)} . \quad (1.2)$$

Для кожного регіону евклідову норму радіус вектора якості життя людей (\bar{C}_{ql}) подамо у такому вигляді:

$$\|\bar{C}_{ql}\| = \sqrt{I_{ec}^2 + I_e^2 + I_s^2} . \quad (1.3)$$

Тоді кількісну міру якості життя людей визначимо як норму проекції цього вектора на ідеальний вектор з координатами (1; 1; 1):

$$C_{ql} = \sqrt{I_{ec}^2 + I_e^2 + I_s^2} \cdot \cos(\alpha) . \quad (1.4)$$

Кут відхилення α радіус вектора \bar{C}_{ql} від ідеального вектора (1,1,1) визначають через значення вимірів I_{ec}, I_e, I_s у такий спосіб:

$$\alpha = \arccos \frac{I_{ec} + I_e + I_s}{\sqrt{3} \cdot \sqrt{I_{ec}^2 + I_e^2 + I_s^2}} , \quad (1.5)$$

$$0 \leq \alpha \leq \arccos \frac{1}{\sqrt{3}} .$$

Отже, відповідно до SDGM метода, норма проекції радіус вектора \bar{C}_{ql} на ідеальний вектор (1,1,1) характеризує якість життя людей, а просторове положення вектора \bar{C}_{ql} у системі координат (I_{ec}, I_e, I_s) характеризує міру «гармонійності» сталого розвитку. Рівновіддаленість вектора \bar{C}_{ql} від кожної з зазначених координат буде відповідати найбільшій гармонійності сталого розвитку, а наближення цього вектора до однієї з координат буде вказувати на пріоритетний розвиток за відповідним виміром і нехтування двома іншими [415].

Величина $G = 1 - \alpha$ є ступенем гармонізації сталого розвитку. Вона зростатиме у разі наближення G до 1 і зменшуватиметься у разі наближення G до 0. Таким чином, компонента якості життя – інтегрована оцінка, яка враховує сумісно усі три виміри сталого розвитку, і, тим самим, відображає взаємозв'язок між трьома нероздільними сферами розвитку суспільства: економічною, екологічною та

соціальною. Ступінь гармонізації сталого розвитку відображає баланс між його економічним, екологічним та соціально-інституціональним вимірами [414 - 415].

Оцінки вимірів сталого розвитку, які використані при визначенні компоненти якості життя повинні ґрунтуватися на даних про широкий спектр явищ різної природи. Також такі оцінки повинні бути інтегрованими, тобто характеризувати певну сторону життєдіяльності людини, як цілісну систему [385].

З метою кількісного оцінювання вимірів сталого розвитку використаємо принципи побудови ієрархічної системи показників та індексів, які визначаються як L_1 -норми:

$$I_i = \sum_{j=1}^n w_j x_{i,j}, i = \overline{1, m}, \sum_{j=1}^n w_j = 1 \quad (1.6)$$

в просторі показників $X^1 \times X^2 \times \dots \times X^m$, що характеризують економічний, екологічний та соціально-інституціональний розвиток кожного i -ого регіону. Вагові коефіцієнти w_j в формулі (1.6) визначаються експертним шляхом.

Використання формули (1.6) потребує узгодження різноманітних даних, як за одиницями виміру, так і за діапазоном значень. Тому, згідно з SDGM методом, якщо більші значення показника x^i відповідають кращому стану сталого розвитку, то застосовується логістичне нормування значень показників за формулою:

$$C_{norm}(x_{i,j}) = \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1}, \quad (1.7)$$

в

я В протилежному випадку, коли найбільші значення показника x^i відповідають гіршому стану сталого розвитку, використовується обернена до обчисленої за формулою (1.7) величина:

й

$$C_{norm}(x_{i,j}) = 1 - \left(1 + e^{\frac{a-x_{i,j}}{b}} \right)^{-1} \quad (1.8)$$

п

а

р

а

Сумарний вплив сукупності загроз на різні регіони України будемо оцінювати за допомогою компоненти безпеки життя людей C_{sl} [414 - 415].

Взаємозв'язки між процесом сталого розвитку, територією, яка знаходиться у фокусі дослідження, та геопросторовими даними можна представити у вигляді системної взаємодії між даними, моделями та методами їх обробки, як представлено на рисунку 1.2.

За оцінками ООН приблизно третина показників, які характеризують процес сталого розвитку не можуть бути отримані з офіційної статистики [74]. Це означає, що для їх розрахунку потрібні інші підходи, які базуються на використанні геопросторових даних та даних дистанційного зондування Землі.

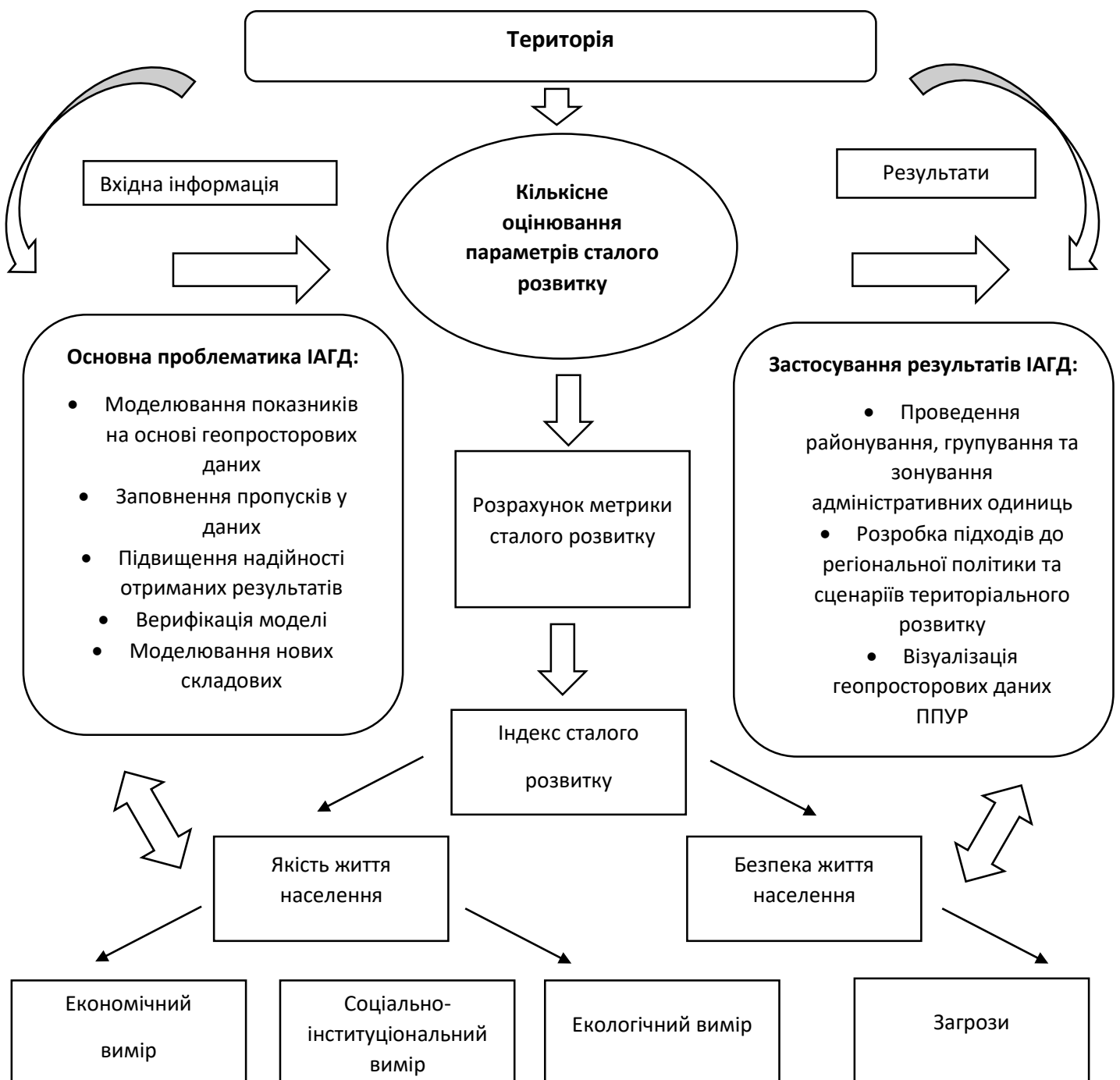


Рис. 1.2. Взаємозв'язки між метрикою сталого розвитку та методами інтелектуального аналізу геопросторових даних (ІАГД)

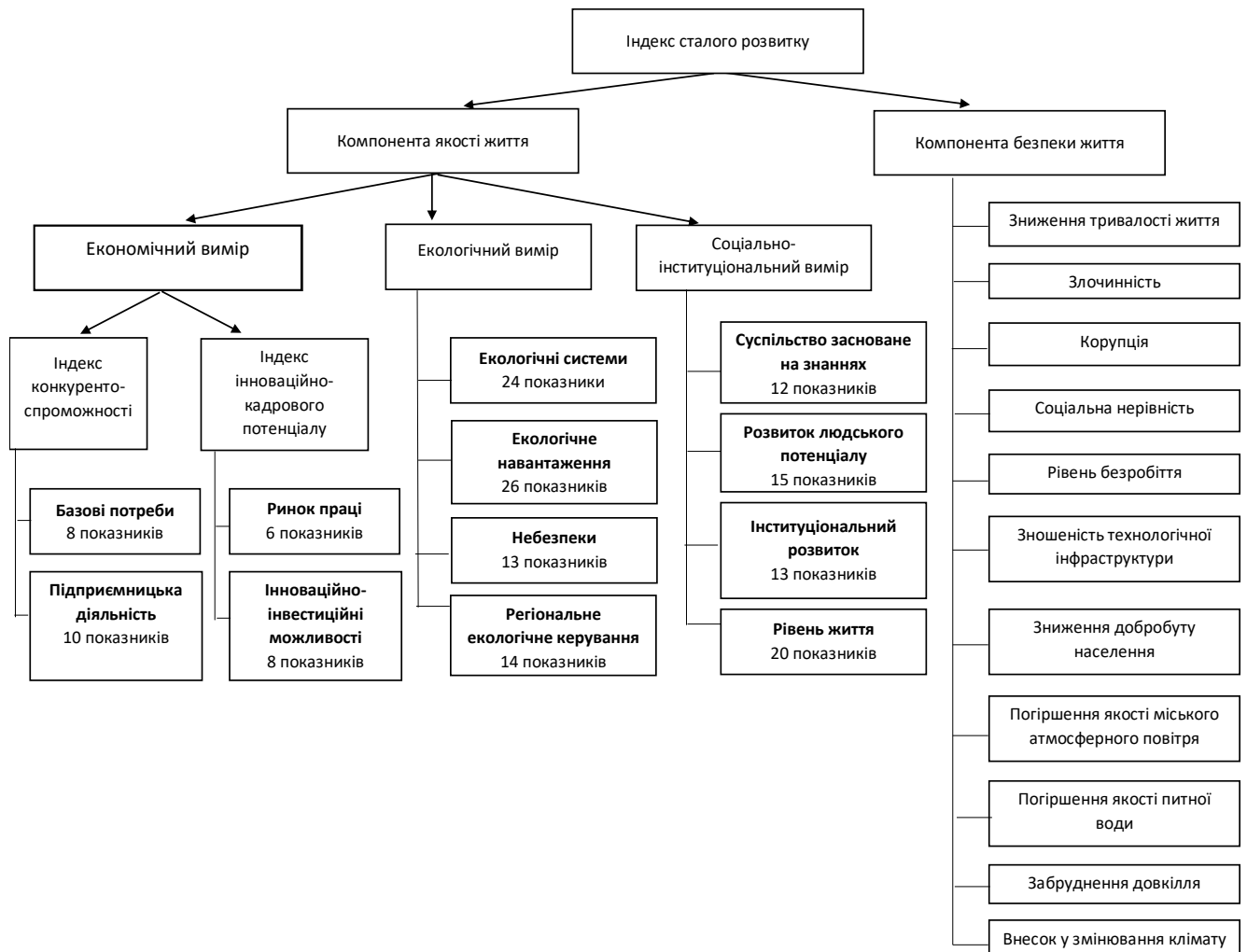


Рис. 1.3. Ієрархічна схема складових показників індексу сталого розвитку

Компонента безпеки життя була розширена за рахунок аналізу загрози «Техногенна небезпека» сталому регіональному розвитку, яка оцінена за допомогою показників кількості техногенних надзвичайних ситуацій в регіонах України та збитків від надзвичайних ситуацій в розрізі регіонів. Ці показники мають значну річну варіабельність за кількістю, обсягом та структурою, тому при аналізі доцільно використовувати усередненні значення за певний період часу. В якості такого періоду пропонується три роки, як часовий проміжок, який показує загальні тенденції у розподілі надзвичайних ситуацій.

Показник кількості техногенних аварій розраховується за формулою:

$$\sum_{i=1}^n E / n,$$

де E – кількість випадків ТНС за рік

n – кількість років, досліджуваного періоду. В даному випадку період дорівнює 3 рокам.

Показник збитків внаслідок ТНС розраховується за формулою:

$$\sum_{i=1}^n S / n,$$

де S – сума збитків від ТНС за рік, тис. грн.

n – кількість років, досліджуваного періоду. В даному випадку період дорівнює 3 рокам.

Геопросторові дані надають можливість для оцінки показників сталого розвитку під час підготовки та збору даних.

В системі взаємозв'язків можна виділити три базові сутності:

- Територія, для якої здійснюється оцінювання сталого розвитку.
- Метрика оцінювання сталого розвитку
- Геопросторові дані про територію та методи їх обробки.

Взаємодія методів ІАГД та метрики оцінювання сталого розвитку полягає у відборі нових та збагачені існуючих показників сталого розвитку на основі геопросторових даних. Тому на першому етапі взаємодія відбувається на рівні вхідних даних системи.

Основною проблематикою застосування ІАГД на етапі підготовки вхідних даних є:

- Моделювання показників на основі геопросторових даних.
- Заповнення пропусків у даних.
- Підвищення надійності отриманих результатів.
- Верифікація моделі.
- Моделювання нових складових показників.

Після обробки даних ІАГД дозволяє отримати додаткові результати за допомогою таких підходів як:

- Проведення районування, групування та зонування адміністративних одиниць.
- Розробка підходів до регіональної політики та сценаріїв територіального розвитку.
- Візуалізація геопросторових даних ППУР.

Сукупність застосування методів ІАГД дозволяє підвищити надійність та точність показників сталого розвитку, запропонувати підходи до їх практичної реалізації.

1.2. Геопросторові дані як об'єкт аналізу

Розвиток людського суспільства завжди відбувався в тісному зв'язку з оточуючим природним середовищем. За останню 1000 років ойкумена охопила усю поверхню Землі і вийшла в космічний простір. Весь цей час модель сприйняття та пізнання всесвіту ґрунтувалась на уявленні про три виміри простору та часову динаміку. Людське пізнання завжди намагалося описати оточуючий простір використовуючи для цього доступні на кожному певному проміжку часу інструменти. В давні часи це були схематичні креслення на камінні, пізніше, з розвитком науки та мови стали розповсюджуватися картографічні зображення, які виконувались різними методами та техніками. В такому вигляді відображення просторової інформації дійшло до середини ХХ сторіччя, коли з'явилися перші геоінформаційні системи і відомості про розміщення геопросторових об'єктів почали зберігатися в базах даних у вигляді запису послідовності координат та їх атрибутивних значень.

Під фізичним простором розуміють тривимірний простір нашого світу, в якому визначається положення фізичних тіл, відбувається механічний рух, геометричне переміщення різних фізичних тіл та об'єктів. Цей простір називається тривимірним, оскільки він має три однорідних виміри - висоту, ширину і довжину,

тобто тривимірний простір описується трьома одиничними ортогональними векторами [54].

В просторі необхідно три декартових координати, які задаються числами (x, y, z) :

x — відстань від точки P до площини yz

y — відстань від точки P до площини xz

z — відстань від точки P до площини xy

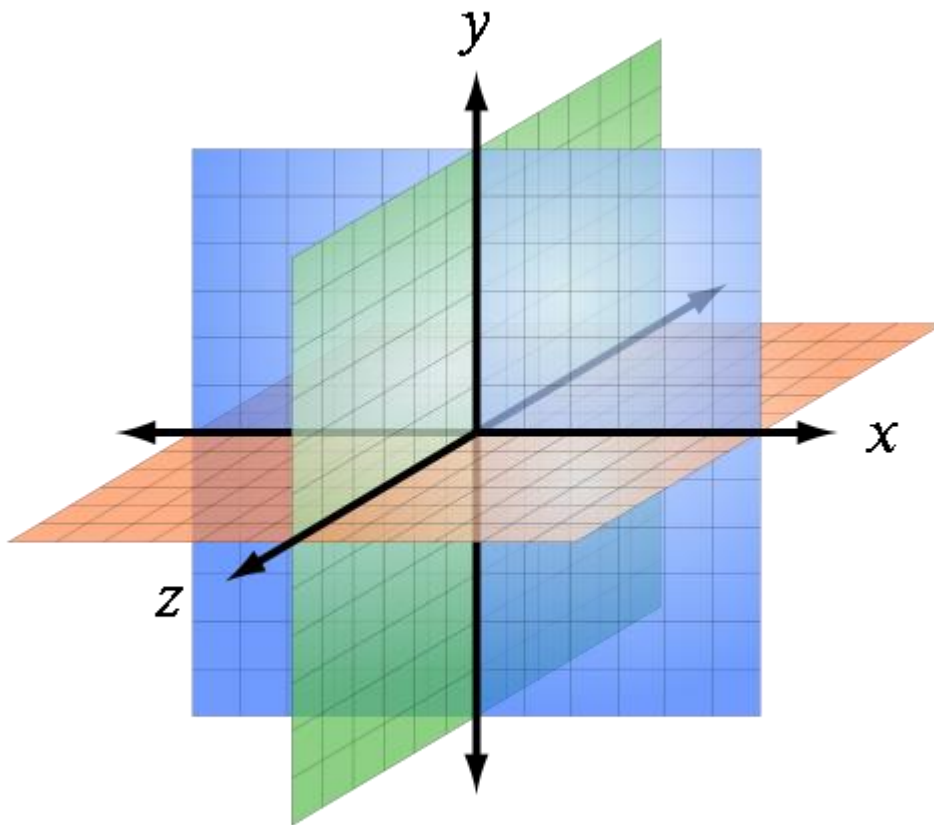


Рис.1.4. Тривимірний фізичний простір

Геопростір є частиною загального фізичного простору, яка обмежена географічною оболонкою Землі та діапазоном фізичних вимірів у масштабах географічних задач.

Геопростір (географічний простір) – «форма існування географічних об'єктів і явищ у межах географічної оболонки; сукупність відносин між географічними об'єктами, розташованими на конкретній території і що розвиваються в часі» [24].

Термін «географічна оболонка» був уведений радянським вченим А. А. Григор'євим для позначення зони взаємодії між основними сферами Землі: літосферою, атмосферою, гідросферою, біосферою та антропосферою [301].

Геопростір характеризується замкнутістю, позитивної кривизною (опуклістю), квазіфрактальністю, анізотропністю в гравітаційному полі Землі. Місцезнаходження в геопросторі визначається сферичними (широта і довгота) або прямокутними координатами, а також щодо інших об'єктів в геопросторі. Анізотропність геопростору виражається в нерівнозначності горизонтального і вертикального напрямків (переміщення по вертикалі значно ускладнено), наслідком чого є сферичність будови Землі, а також (меншою мірою) широтного і довготного напрямків, проявом чого є широтна зональність. Мірами відстані в географічному просторі може бути не тільки Декартова відстань, а і час, енергія або кошти на подолання відстані [301].

Час виступає додатковим четвертим виміром існування об'єктів у просторі. Тому важливо розглядати просторово-часовий аспект геопросторового об'єкту.

Простір-час (просторово-часовий континуум) – фізична модель, яка доповнює простір рівноправним [291] тимчасовим виміром і таким чином створює теоретико-фізичну конструкцію, яка називається просторово-часовим континуумом. Простір-час є безперервним і з математичної точки зору являє собою різноманіття з Лоренцевою метрикою.

У нерелятивістської класичній механіці доречно використання Евклидова простору, що не залежний від одновимірного часу, замість простору-часу, оскільки час розглядається як загальне і незмінне, будучи незалежним від стану руху спостерігача. У разі релятивістських моделей час не може бути відокремлений від трьох вимірів простору, тому що спостерігаєма швидкість, з якою тече час для об'єкта, залежить від його швидкості відносно спостерігача, а також від сили гравітаційного поля, яке може уповільнити перебіг часу [216].

На Землі широта і довгота – це дві незалежні координати, які разом однозначно визначають положення. У просторі-часі координатна сітка, яка простягається в $3 + 1$ вимірах, локалізує події (замість просто точки в просторі), тобто час додається як ще один вимір в координатній сітці. Таким чином, координати визначають де і коли відбуваються події. Однак єдина природа простору-часу і його незалежність від вибору координат дозволяють припустити, що щоб визначити тимчасову координату в одній системі координат, необхідні як тимчасова, так і просторова координати в іншій системі координат. На відміну від звичайних просторових координат, в просторі-часі виникає поняття світлового конуса, що накладає обмеження на допустимі координати, якщо одна з них скрізь повинна бути тимчасовою. Ці обмеження жорстко пов'язані з особливою математичною моделлю, яка відрізняється від Евклидова простору з його очевидною симетрією.

Дані — це інформація, подана у формалізованому вигляді, прийнятному для обробки автоматичними засобами за можливої участі людини [49].

Геопросторові дані — це дані про об'єкти та явища, які безпосередньо або опосередковано пов'язані з місцеположенням на Землі, що визначені у певній системі просторово-часових координат [203].

Звідси можна зробити висновки про основні параметри геопросторових даних:

- Вимірність.
- Координованість.
- Топологічність.
- Семантичність.
- Координованість даних лежить в основі визначення їх положення в реальному геопросторі.

Для цього використовуються загальноприйняті системи координат, більшість із яких має в своїй основі або Декартову систему координат, в якій вісі вимірів є ортогональними між собою.

Тривимірні прямокутні координати (X,Y,Z)

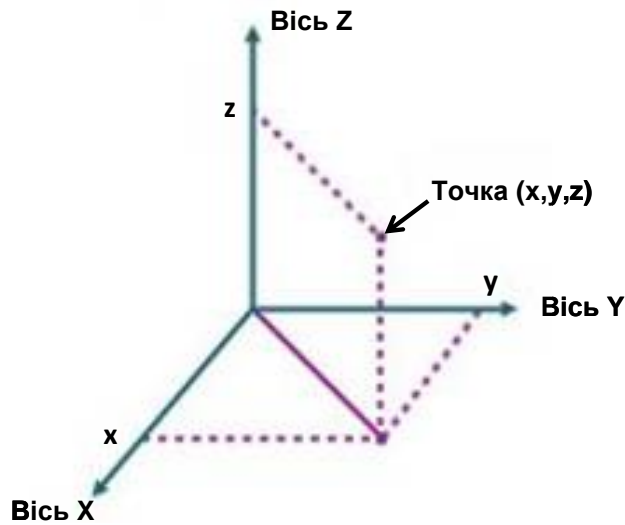


Рисунок 1.5. Тривимірні прямокутні координати (X,Y,Z)

Для розрахунку відстані D між точками (X_a, Y_a, Z_a) та (X_b, Y_b, Z_b) у прямокутній системі у тривимірному просторі використовується рівняння:

$$D = \sqrt{(X_b - X_a)^2 + (Y_b - Y_a)^2 + (Z_b - Z_a)^2}. \quad (1.9)$$

Інший шлях визначення координат місця розташування у дво- або тривимірному просторі полягає у використанні **полярної системи координат** (або радіальної системи координат).

Полярні координати засновані на кутах та радіальних відстанях. У двовимірній полярній системі координат ми маємо один кут та радіус (рис. 1.5), а у тривимірній полярній системі координат – маємо два кути та радіус. Визначення одного або двох кутів та радіусу дає можливість побудувати вектор, що встановлює місце розташування відносно початкової точки. У випадку географічних координат початковою точкою є центр Землі, першим кутом є широта, а другим кутом – довгота. Радіальна відстань є загально прийнятою, оскільки ми вважаємо, що місце розташування має бути на поверхні Землі.

Двовимірні полярні координати (r, θ)



Рис. 1.6. Двовимірні полярні координати [87]

Якобіан перетворення полярних координат в Декартові матиме вигляд:

$$\det \frac{\partial(x_1, \dots, x_n)}{\partial(r, \varphi, \vartheta_1, \dots, \vartheta_{n-2})} = r^{n-1} \sin \vartheta_1 (\sin \vartheta_2)^2 \dots (\sin \vartheta_{n-2})^{n-2}, \quad (1.10)$$

де n -вимірний елемент об'єму матиме вигляд:

$$\begin{aligned} dV &= r^{n-1} \sin \vartheta_1 (\sin \vartheta_2)^2 \dots (\sin \vartheta_{n-2})^{n-2} dr d\varphi d\vartheta_1 \dots d\vartheta_{n-2} \\ &= r^{n-1} dr d\varphi \prod_{j=1}^{n-2} (\sin \vartheta_j)^j d\vartheta_j \end{aligned} \quad (1.11)$$

Ці види базових систем координат застосовуються для опису координат геопросторових об'єктів на поверхні Землі. Оскільки форма Землі є неправильною, то для її опису застосовуються аналітичні математичні конструкції: сфероїди та еліпсоїди.

Форма Землі описується формою геоїда, під яким розуміється поверхня постійного гравітаційного потенціалу (рис. 1.7).

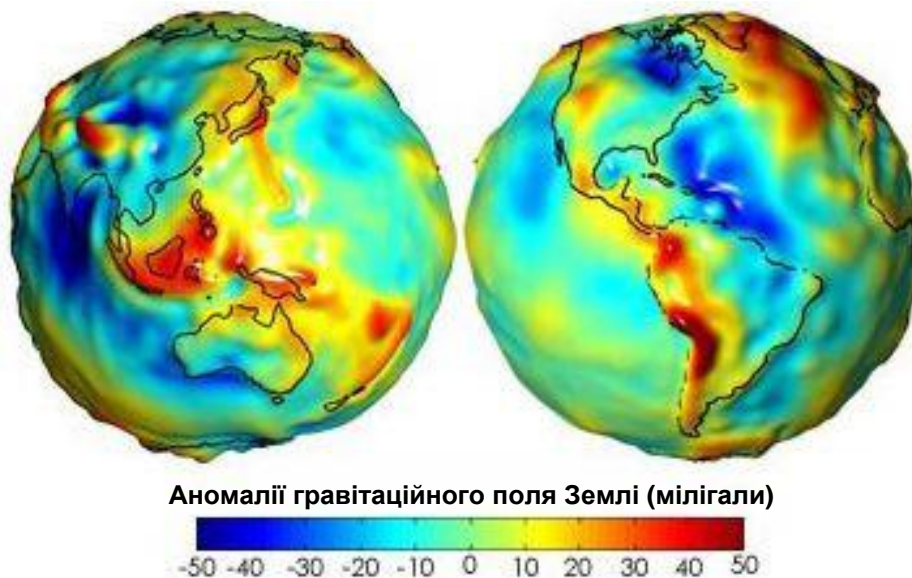


Рис. 1.7. Геоїд, поверхня постійного гравітаційного потенціалу, є досить змінною за формою. Створено на основі даних місії GRACE [278]

1. Географічна система координат

Системою географічних координат називається система, у якій розміщення точки на земній поверхні визначається кутовими величинами (широтою та довготою) відносно площини екватора та початкового меридіана [26]. В основі географічної системи координат лежить сфероїд, який приблизно описує форму Землі (рис. 1.8).



Рисунок 1.8. Елементи географічної системи координат

Концепція широти і довготи (ϕ , λ) і вертикальна відстань (висота) над поверхнею геоїда (H) є основою для будь-якого опису географічних координат на

поверхні Землі. Висота (H) визначається за допомогою геоїда поверхні постійного гравітаційного потенціалу.

2. Геодезична система координат

Географічні координати є генералізованою концепцією геодезичних координат. В геодезичних координатах, поверхня Землі апроксимується еліпсоїдом, і розташування на поверхні еліпсоїда описується в термінах геодезичної широти (B) і довготи (L). Геодезична широта точки на поверхні еліпсоїда є кутом між екваторіальною площиною і нормаллю еліпсоїда. Довгота точки на поверхні еліпсоїда є кутом на схід чи захід від опорного меридіана до іншого меридіану, який проходить через цю точку. Всі меридіани є половинками великих еліпсів, які сходяться на північному і південному полюсах.

3. Глобальна геоцентрична прямокутна система координат (X, Y, Z)

Глобальні прямокутні координати (X, Y, Z) є глобальною системою координат для всієї Землі. Відправною точкою або початком відліку цієї системи координат, як правило, є центр маси Землі.

З початком відліку в центрі еліпсоїда, Z -вісь орієнтована уздовж осі еліпсоїда, позитивна на північ, X і Y -осі орієнтовані в площині екватора, значення осі X позитивні до 0 градусів довготи, і Y -осі значення позитивні до 90 градусів східної довготи (рис 1.9).

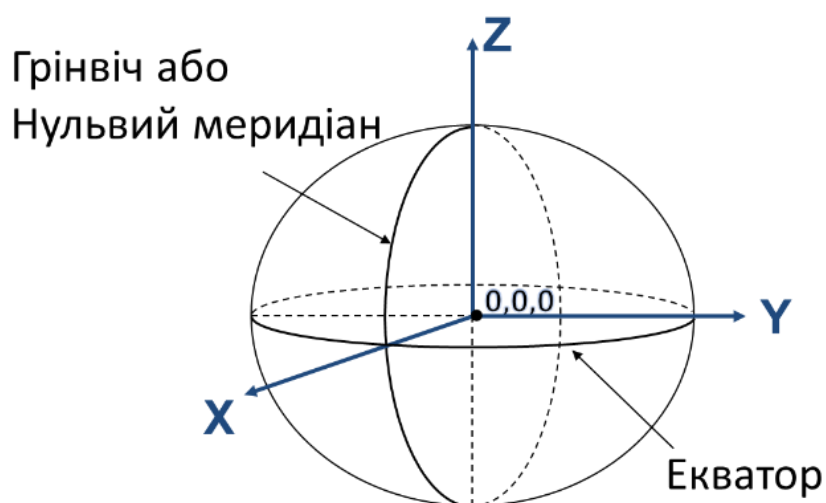


Рисунок 1.9. Глобальна геоцентрична прямокутна система координат

Об'єктом інформаційного моделювання в ГІС є просторовий об'єкт. Це одне з ключових понять геоінформатики. Він може бути визначений як цифрове представлення (модель) об'єкта реальності (місцевості), що містить його місцеположення і набір властивостей (характеристик, атрибутів), або сам цей об'єкт. Певна множина цифрових даних про просторові об'єкти утворює просторові дані. Вони складаються з двох взаємопов'язаних частин: позиційної (Тополого-геометрична) і непозиційна (атрибутивна) складових, які створюють опис просторового положення і тематичного змісту даних [26].

Фіксація моделі еліпсоїду відносно геоїда створює **датум**, який визначає початкову точку, прив'язану до поверхні або центру Землі (горизонтальний датум) та визначає базову поверхню, від якої рахуються висоти (вертикальний датум).

Для перенесення даних з поверхні еліпсоїда на площину використовуються проєкційні системи координат, в основі яких лежить механізм переходу між сферичними та плоскими координатами. За рахунок цього усі проєкції мають певні спотворення відносно площі, форми та довжин відображених об'єктів. За рахунок використання геоінформаційних систем можливе збереження даних у географічній проєкції координат без спотворення на площині. В той же час для публікації необхідно застосовувати картографічних даних проєкції.

Таким чином, опис місцеположення об'єктів в геопросторі містить вибір системи координат, певного датуму та проєкції для відображення. Перетворення між різними видами проєкції відбуваються в автоматичному порядку у разі, коли усі ці параметри задані для геоданих.

В певному сенсі геопростір і відповідні системи відліку в ньому виступають основою для інтеграції різноманітної за способами та часом створення інформації, що посилює роль ГІС для сумісного аналізу даних.

Опис просторових об'єктів ґрунтується на використанні математичних моделей опису їх властивостей.

Просторові об'єкти як абстрактні уявлення реальних об'єктів і предмет інформаційного моделювання (цифрового опису) в ГІС різноманітні і традиційно класифікуються відповідно до характеру просторової локалізації відображуваних

ними об'єктів реальності, мірності простору, який вони утворюють, моделі даних, яка використовується для їх опису та ін. В рамках об'єктно-орієнтованих моделей даних можуть конструюватися нові класи об'єктів, відмінні від базових або створених раніше [362].

Базовими (елементарними) типами просторових об'єктів, якими оперують сучасні ГІС зазвичай вважаються:

- точка (точковий об'єкт) - 0-мірний об'єкт, який характеризується плановими координатами;
- лінія (лінійний об'єкт, полілінія) - 1-мірний об'єкт, створений послідовністю не менше двох точок з відомими плановими координатами (лінійними сегментами або дугами);
- полігон - 2-мірний (площинний) об'єкт, внутрішня область, обмежена замкнутою послідовністю ліній (дуг в векторних топологічних моделях даних або сегментів в моделі «Спагеті») та ідентифікована внутрішньою точкою (міткою);
- піксель (піксель) - 2-мірний об'єкт, елемент цифрового зображення, найменша з його складових, що отримується в результаті дискретизації зображення (розбиття на далі неділимі елементи растра), елемент дискретизації координатної площини в растровій моделі (даних) ГІС;
- комірка (регулярна комірка) - 2-вимірний об'єкт, елемент розбиття земної поверхні лініями регулярної мережі;
- поверхня (рельєф) - 2-вимірний об'єкт, який визначається не тільки плановими координатами, а й атрибутом Z , який входить до числа атрибутів, що утворюють її об'єкти; оболонка тіла;

тіло - тривимірний (об'ємний) об'єкт, який описується трійкою (триплетом) координат, включаючи аплікату Z , і обмежений поверхнями [26].

Загальний цифровий опис просторового об'єкта включає:

- Найменування;
- Опис місця розташування (місцезнаходження, локалізації);
- Набір властивостей;

- Відносини з іншими об'єктами;
- Просторова «поведінка» [27].

Найменуванням об'єкта служить його географічне найменування (ім'я власне, якщо воно є), його умовний код та / або ідентифікатор, що привласнюється користувачем або призначуваний системою.

Залежно від типу об'єкта його місцеположення визначається парою (триплетом) координат (для точкового об'єкта) або набором координат, організованим певним чином у рамках деякої моделі даних.

Це геометрична частина опису даних, геометрія (метрика) описуваних просторових об'єктів, відмінна від їх семантики (непозиційних властивостей).

Перелік властивостей відповідає атрибутам об'єкта, якісним і кількісним його характеристикам, які привласнюються йому в цифровому вигляді користувачем та можуть бути отримані в ході обробки даних або генеруються системою автоматично (до останнього типу атрибутів належать, наприклад, значення площ і периметрів полігональних об'єктів).

Під атрибутами розуміються саме змістовні, тематичні (непозиційні, непросторові) атрибути (властивості) об'єктів.

Під відносинами розуміють насамперед топологічні відносини (топологію). До топологічних властивостей просторових об'єктів прийнято відносити його розмірність (мірність, просторову розмірність), згідно якої вище були виділені нуль-, одно-, дво- і тривимірні об'єкти; замкнутість.

Якщо мова йде про лінійні об'єкти в широкому сенсі слова: зв'язність; простота (відсутність самонакладання лінійних об'єктів і «островів» в полігоні); знаходження на кордоні, всередині або поза полігоном; ознака точкового об'єкта, яка вказує, чи є він кінцевим для деякої лінії.

Топологія разом з геометрією утворює тополого-геометричну частину опису даних, його позиційну частину.

Таким чином, у найзагальнішому вигляді в просторових даних слід розрізняти і виділяти три складові частини: топологічну, геометричну і

атрибутивну - «геометрію», «топологию» і «атрибутику» цифрової моделі просторового об'єкта.

На концептуальному рівні вся множина моделей просторових даних може бути розділена на три типи: моделі дискретних об'єктів, моделі безперервних полів і моделі мереж [98].

Основним поняттям растрової моделі є растр – прямокутна решітка, яка розбиває зображення на однорідні складові, що називаються пікселями, кожному з яких відповідає певний код, що зазвичай ідентифікує колір в тій чи іншій системі кольорів (колірній моделі). З множини значень логічних пікселів складається цифрове зображення. Растрова модель даних в ГІС припускає розбивку простору (координатної площини) її просторовими об'єктами на аналогічні пікселю дискретні елементи, впорядковані у вигляді прямокутної матриці. Для цифрового опису позиціонування точкового об'єкта при цьому буде достатньо вказати його приналежність до того чи іншого елементу дискретизації.

Простота машинної реалізації операцій з растровими даними знаходиться в протиріччі з іншою головною їх особливістю - значними витратами машинної пам'яті, необхідної для їх зберігання (у порівнянні з обсягами даних в векторних моделях). Існують способи стиснення, компресії растрових даних. Найпростіший з них - групове кодування. Груповий код конвертує вихідний растровий шар в ряд пар цілих чисел, непарні позиції якого відводяться для вказання числа повторюваних пікселів (груп) зі значеннями, що займають парні позиції ряду, утворюючи лічильник і значення групи відповідно.

Узагальнений клас векторних моделей включає два їх типи: векторні топологічні і векторні нетопологічні моделі. Вони використовуються для цифрового відображення точкових, лінійних і полігональних об'єктів [117].

В сучасних геоінформаційних системах геопросторові дані описуються 4 виміром – час (t). Це дозволяє моделювати зміну об'єкту, його місцеположення, топології та геометрії у часі, що надає нові можливості для інтелектуального аналізу даних. Особливий розвиток аналізу динамічних процесів відбувся за рахунок використання технологій GNSS.

1.3. Джерела, методи збору та інтеграції геопросторових даних

Джерелами геопросторових даних можуть бути будь-які інформаційні ресурси та бази даних, що містять відомості про просторове розташування. За розрахунками дослідників близько 80 % інформації, яка обробляється людиною, містить просторові дані або прив'язку до місцевості. Звичайно, частка геопросторової інформації, що обробляється на сьогодні становить менше 80%, але вона постійно збільшується під впливом розвитку інформаційних технологій. Особливо швидке зростання розпочалось із розвитком соціальних мереж та сервісів, які стали новим джерелом геопросторових даних. Бурхливого розвитку поняття «джерела геопросторових даних» зазнало під впливом розвитку концепції глобальних систем позиціонування, сенсорів та інтернету речей.

Ведення даних є обов'язковою операцією, необхідною для функціонування ГІС. Для різних типів даних розроблені спеціальні **технології збору та введення**, що відповідають функціональним можливостям, включеним до складу програмного ГІС-забезпечення та розроблені спеціалізовані периферійні пристрої [27].

Як вихідні матеріали, з яких виконується введення даних у ГІС, у наш час використовуються:

- топографічні карти;
- загальногеографічні карти різного тематичного змісту;
- архітектурні плани і плани землевпорядкування;
- дані дистанційного зондування Землі (ДЗЗ);
- матеріали польової інструментальної зйомки;
- стандартні статистичні звітні форми в паперовому та електронному поданні;
- текстові джерела, фотографії та ілюстрації;
- рукописні карти і тексти [27].

Геоінформаційні системи оперують різними впорядкованими наборами даних. Серед них традиційно розрізняють картографічні, статистичні, аерокосмічні матеріали, які трансформуються і вводяться в середовище ГІС та / або

запозичуються з інших геоінформаційних систем. Крім зазначених матеріалів, набагато рідше використовуються дані спеціально проведених польових досліджень і зйомок, а також літературні (текстові) джерела, що дає нам право охарактеризувати їх лише в самому загальному вигляді. «Тип джерела» об'єднує генетично сильно розрізнені за комплексом характеристик ознаки, що і буде проаналізовано нижче. До них належить, наприклад, така важлива ознака: в якій цифровій (векторній, растровій) або нецифровій (аналоговій) формі надходить, зберігається і використовується той чи інший набір даних, від чого залежать легкість, вартість і точність введення цих даних в цифровому середовищі ГІС [213].

Під час аналізу та оцінки різних типів джерел як основи інформаційного забезпечення ГІС слід мати на увазі їх загальні властивості: просторове охоплення, масштаби, розрізнення, якість, форму існування (аналогова – цифрова), періодичність надходження, актуальність і оновлюваність, умови та вартість отримання, придбання і переведення в цифрову форму (цифрування), доступність, формати представлення, відповідність стандартам і інші характеристики, які об'єднуються узагальнюючим терміном «метадані».

Загалом, джерела просторових даних можна розділити на структуровані, частково структуровані та неструктуровані дані. Інша класифікація джерел спирається на способи отримання даних: первинні, вторинні та імпортовані.

Основним джерелом даних в ГІС є шари геопросторових даних та геопросторові бази геоданих. Ці дані продукуються в ГІС та інших інформаційних системах і можуть в подальшому використовуватись в ГІС як вхідне джерело даних. Часто такі дані отримують назву базових наборів геопросторових даних, які становлять основу розвитку інфраструктури геопросторових даних. Цей вид даних можна віднести до імпортованих структурованих даних.

Первинні дані надходять до ГІС від сенсорів та за результатами польових безпосередніх або дистанційних зйомок.

Вторинні дані отримуються шляхом конвертації в ГІС формат вже існуючих джерел даних. До таких джерел належать вже виготовлені картографічні матеріали, інженерні плани, паперові дані, непросторові бази даних. В залежності від

поставлених завдань ці матеріали конвертуються і відповідно обробляються для використання в ГІС. Стандартна процедура конвертації паперових даних складається з:

- Сканування паперової карти;
- Прив'язка сканованого зображення у відповідній системі координат;
- Векторизація або дигіталізація растрового сканованого зображення;
- Створення атрибутів отриманих векторних об'єктів.

Ці дані створили основу сучасних геопросторових даних. Оскільки вони отримувались на основі стандартизованих картографічних даних, то могли легко бути описані метаданими та відповідати єдиним вимогам. Однак на даний час ситуація кардинально змінилась і на перше місце вийшли дані обробки даних дистанційного знімання Землі та дані, які надходять від сенсорів [306].

Дистанційне зондування Землі (ДЗЗ) - спостереження поверхні Землі наземними, авіаційними і космічними засобами, які оснащено різними видами знімальної апаратури. Робочий діапазон довжин хвиль, що приймаються знімальною апаратурою, становить від часток мікрометра (видиме оптичне випромінювання) до метрів (радіохвилі). Методи зондування можуть бути пасивні, тобто використовують природне відбите або вторинне теплове випромінювання об'єктів на поверхні Землі, обумовлене сонячною активністю, і активні - використовують вимушене випромінювання об'єктів, ініційоване штучним джерелом направленої дії. Дані ДЗЗ, отримані з космічного апарату (КА), характеризуються великим ступенем залежності від прозорості атмосфери. Тому на КА використовується багатоканальне обладнання пасивного і активного типів, що реєструє електромагнітне випромінювання в різних діапазонах [29].

Апаратура ДЗЗ перших КА, запущених в 1960-70-х рр. була трасового типу - проекція області вимірювань на поверхню Землі представляла собою лінію. Пізніше з'явилася і широко поширилася апаратура ДЗЗ панорамного типу - сканери, які знімають проекцію області вимірювань на поверхню Землі у вигляді смуги [29].

Космічні апарати для дослідження природних ресурсів оснащуються в основному оптичної або радіолокаційної апаратурою. Переваги останньої

полягають в тому, що вона дозволяє спостерігати поверхню Землі в будь-який час доби, незалежно від стану атмосфери [27].

Дистанційне зондування є методом отримання інформації про об'єкт або явище без безпосереднього фізичного контакту з даним об'єктом. У сучасному розумінні, термін в основному відноситься до технологій повітряного або космічного зондування місцевості з метою виявлення, класифікації та аналізу об'єктів земної поверхні, а також атмосфери і океану, за допомогою розповсюдження сигналів (наприклад, електромагнітного випромінювання). Поділяють на активну (сигнал спочатку випромінюється літаком або космічним супутником) і пасивне дистанційне зондування (реєструється тільки сигнал інших джерел, наприклад, сонячне світло).

Пасивні сенсори дистанційного зондування реєструють сигнал, що випромінюється або відбитий об'єктом або прилеглою територією. Відбите сонячне світло - найбільш часто використовуване джерело випромінювання, що реєструється пасивними сенсорами.

Активні прилади, в свою чергу, випромінюють сигнал з метою сканування об'єкта і простору, після чого сенсор має можливість виявити і виміряти випромінювання, відбите або утворене шляхом зворотного розсіювання з метою зондування. Прикладами активних сенсорів дистанційного зондування є радар і лідар, якими вимірюється затримка в часі між випромінюванням і реєстрацією поверненого сигналу, таким чином визначаючи розміщення, швидкість і напрямок руху об'єкта.

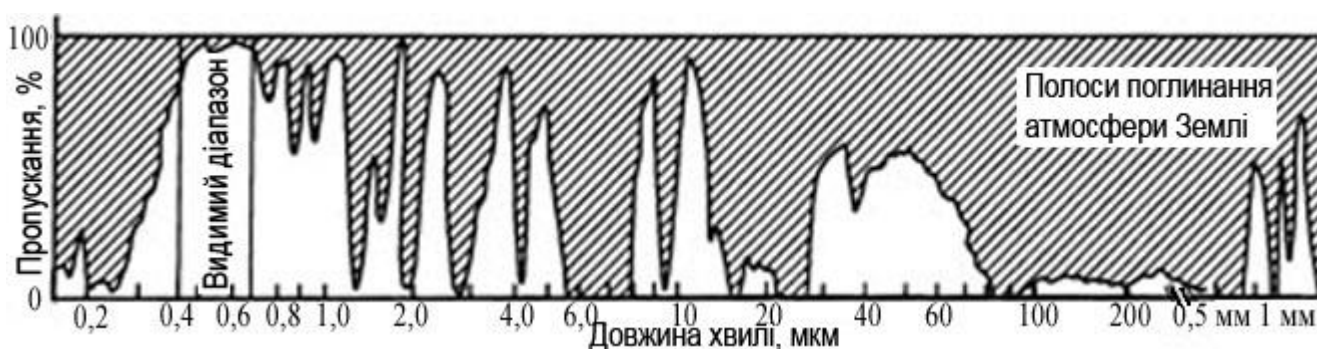


Рис.1.10. Просторова роздільна здатність в ДЗЗ: просторова, радіометрична, спектральна, часова [54]

В даний час для зйомок з космосу найбільш часто використовуються багатоспектральні оптико-механічні системи - сканери, встановлені на ШСЗ різного призначення. За допомогою сканерів формуються зображення, що складаються з безлічі окремих, послідовно одержуваних елементів. Термін «сканування» позначає розгортку зображення за допомогою скануючого елемента, який поелементно переглядає місцевість поперек руху носія і посиляє променистий потік в об'єктив і далі на точковий датчик, що перетворює світловий сигнал в електричний. Цей електричний сигнал надходить на приймальної станції по каналах зв'язку. Сканерне зображення можна отримати у всіх спектральних діапазонах, але особливо ефективним є видимий та ІЧ-діапазони. При зйомці земної поверхні за допомогою скануючих систем формується зображення, кожному елементу якого відповідає яскравість випромінення ділянки, що знаходиться в межах миттєвого поля зору. Сканерне зображення - упорядкований пакет яскравості даних, переданих по радіоканалах на Землю, які фіксуються в цифровому вигляді і потім можуть бути перетворені в кадрову форму [30].

Найважливішою характеристикою сканера є кут сканування (огляду) і миттєвий кут зору, від величини якого залежать ширина смуги знімання і роздільна здатність. Залежно від величини цих кутів сканери ділять на точні і оглядові. У точних сканерів кут сканування зменшують до $\pm 5^\circ$, а у оглядових збільшують до $\pm 50^\circ$. Величина роздільної здатності при цьому обернено пропорційна ширині смуги, що знімається [30].

Таблиця 1.1. Характеристики спектральних діапазонів [229]

Область спектру	Ширина області спектру
Видима область, мкм	
<i>Кольорові зони</i>	
фіолетова	0.39-0.45
синя	0.45-0.48
блакитна	0.48-0.51
зелена	0.51-0.55

жовто-зелена	0.55-0.575
жовта	0.575-0.585
помаранчова	0.585-0.62
червона	0.62-0.80
Область ІЧ випромінення, мкм	
ближня	0.8-1.5
середня	1.5-3.0
дальня	>3.0
Радіохвильова область, см	
X	2.4-3.8
C	3.8-7.6
L	15-30
P	30-100

Добре зарекомендував себе сканер нового покоління, названий «тематичним картографом», яким були оснащені американські ШСЗ Landsat 7 і Landsat 8. Сканер типу «тематичний картограф» працює в семи діапазонах з роздільною здатністю 30 м у видимому діапазоні спектра та 120 м в ІЧ-діапазоні [31].

Сучасний ринок даних ДЗЗ надає широкі можливості щодо вибору КС, які можуть бути використані для створення та оновлення геопросторової інформації. Для завдань великомасштабного топографічного картографування підійдуть знімки надвисокої роздільної здатності (вище 2 м), такі як WorldView-1, GeoEye-1, CARTOSAT-2, QuickBird, IKONOS, РесурсДК та ін., і високої роздільної здатності (2-3 м), такі як Radarsat-2, SPOT-5, CARTOSAT-1, FORMOSAT-2, ALOS PRISM.

Таблиця 1.2. Співвідношення масштабу карт з просторовою роздільною здатністю знімків [367]

Датчик	Розмір пікселя	Можливий масштаб
TERRA (ASTER) TIR	90 м	1:1 000 000

TERRA (ASTER) SWIR	30 м	1:300 000
TERRA (ASTER) VNIR	15 м	1:100 000
RapidEYE	5 м	1:25 000
ALOS AVNIR	10 м	1:100 000
ALOS PRISM	2.5 м	1:12 500
EROS	0.7 м	1:5 000
IKONOS MSI	3,2 м	1:20 000
IKONOS PAN та PSM	0,82 м	1:5 000
QUICKBIRD MSI	2.44 м	1:12 500
QUICKBIRD PAN та PSM	0.61 м	1:2 000
WorldView-2 MSI	1,84 м	1:10 000
WorldView-2 PAN та PSM	0,46 м	1:2 000
WorldView-1	0.5 м	1:2 000
GeoEYE MSI	1,65 м	1:10 000
GeoEYE PAN та PSM	0.41 м	1:2 000

Для цілей тематичного картографування, в більшості випадків, підходять КС сенсорів середнього просторового розрізнення, які мають велику кількість спектральних каналів, що відповідають за вузькі ділянки спектра. Наявність каналів в середньому, ближньому і далекому ІЧ діапазонах дає розширені можливості для прикладних задач, пов'язаних з картографування рослинного та ґрунтового покривів. До таких знімальних систем відносяться, насамперед, RapidEye, ALOS AVNIR-2, Landsat-7/8, SVOT-4/5, ASTER, IRS-1C / ID, Resourcesat-1 та ін.

Сучасні КС нижчої роздільної здатності в таких системах, як MODIS, MERIS, NOAA та інших, також несуть величезний обсяг корисної інформації, що дозволяє вирішувати тематичні завдання в дрібних масштабах.

До теперішнього моменту накопичено великий архів даних ДЗЗ, який регулярно поповнюється. У потенційного споживача є широкі можливості вибору знімків за типом зйомки, просторового і радіометричного розрізнення, а також за часом зйомки. Архівні дані, як правило, значно дешевше оперативних, але в будь-якому випадку, для кінцевого користувача може минути деякий час між замовленням знімка і його отриманням, навіть якщо знімки купуються через Інтернет. Організації, які працюють з великими обсягами інформації, можуть отримувати дані з супутників в режимі оперативного зв'язку, минаючи постачальників даних ДЗЗ, якщо придбають і встановлять у себе станцію прийому інформації.

GPS (англ. Global Positioning System - система глобального позиціонування) - супутникова система навігації, яка забезпечує вимірювання відстані, часу і визначає місце розташування у всесвітній системі координат WGS 84. Дозволяє майже при будь-якій погоді визначати місце розташування в будь-якому місці Землі (виключаючи приполярні області) і навколоземного космічного простору. Система розроблена, реалізована і експлуатується Міністерством оборони США, при цьому в даний час доступна для використання для цивільних цілей - потрібен тільки навігатор або інший апарат (наприклад, смартфон) з GPS-приймачем.

Основний принцип використання системи - визначення місця розташування шляхом вимірювання моментів часу прийому синхронізованого сигналу від навігаційних супутників антеною споживача. Для визначення тривимірних координат GPS-приймача потрібно мати чотири рівняння: «відстань дорівнює добутку швидкості світла на різницю моментів прийому сигналу споживачем і моменту його синхронного випромінювання від супутників» [29].

GPS складається з трьох основних сегментів: космічного, керуючого і призначеного для користувача. Супутники GPS транслюють сигнал з космосу, і все приймачі GPS використовують цей сигнал для обчислення свого положення в просторі за трьома координатам в режимі реального часу. Космічний сегмент складається з 32 супутників, що обертаються на середній орбіті Землі [29].

Геотегінг - процес приєднання географічних метаданих до різних інформаційних ресурсів, таких як веб-сайти, RSS, фотографії, що будь-яким чином характеризують ці ресурси (точка зйомки, місце розташування і т. п.). Ця інформація складається, як правило, з координат широти і довготи, хоча вона може включати також висоту, відстань і напрям до населених пунктів. Геотегінг може допомогти в пошуках різної інформації, прив'язаної до певної місцевості. Наприклад, можна проводити пошук фотографій, що зроблені поблизу від певного місця шляхом введення координат в пошукову систему з підтримкою геотегінгу. Пошукові системи з підтримкою геотегінгу можуть бути корисними для пошуку прив'язаних до певного місця новин, веб-сайтів, або інших ресурсів. У переважній більшості випадків в побуті використовується геотегінг фотографій [390].

IP геолокації – це метод для визначення IP адреси суб'єкта, який потім визначає, яка країна, штат, місто, поштовий індекс, організація або місце розташування було пов'язане з IP-адресою. Є кілька провайдерів IP геолокації в галузі постачання IP бази даних адрес. Ці бази даних містять дані IP-адреси, які можуть бути використані в брандмауері, сервері доменних імен, оголошених серверах, маршрутизації, поштовій системі, веб-сайтах та інших автоматизованих системах, де геолокація може бути корисною [381].

W3C Geolocation API представляє собою спробу World Wide Web Consortium (W3C), щодо стандартизації інтерфейсу для отримання географічної інформації місця для клієнтських пристроїв. Найбільш поширеними джерелами інформації про місцезнаходження є IP-адреси, Wi-Fi і Bluetooth MAC адреси, мітки радіочастотної ідентифікації (RFID) або пристрої глобальної системи позиціонування (GPS) та GSM / CDMA ідентифікатори осередків. Інформація про розташування повертається із заданою точністю в залежності від доступного джерела інформації про розташування [236].

Геокодування часто пов'язано з геолокацією. Це процес пошуку пов'язаних географічних координат (широти і довготи) з інших географічних даних, таких як місце розташування або адреса. Зворотне геокодування є знаходженням пов'язаних текстових місці, наприклад, адрес, географічних координат [366].

Для цілей управління територіями необхідно мати можливість локалізувати геопросторові дані. З цією метою створені системи, де застосовується поняття *інфраструктура просторових даних (ІПД)*. Воно широко використовується на різних рівнях управління, у тому числі державному, а також в географії, картографії, геоінформатиці і в інших дисциплінах. Інфраструктуру просторових даних можна визначити як систему, що забезпечує взаємодію осіб, які використовують у своїй роботі різні цифрові просторові дані, а також як єдине середовище, що забезпечує пошук, публікацію і обмін різноманітними геоінформаційними ресурсами. ІПД - ієрархічна впорядкована система, побудована з використанням кібертехнологій та геоінформаційних підсистем, заснована на загальних просторових даних і метаданих, а також мережі географічних інформаційних вузлів - геопорталів і каталогів метаданих) [85].

Незважаючи на те, що в роботі можна безпосередньо користуватися цифровими просторовими даними, ІПД значно розширює управлінські можливостями. Проблема полягає в тому, що обсяг таких геопросторових даних швидко росте і доступні джерела інформації (Інтернет) переповнені «інформаційним шумом». Створення ІПД допомагає вирішити цю проблему за рахунок формування особливого пошукового і обмінного середовища зі специфічними інструментами пошуку інформації за стандартизованими описами (метаданими), а також по карті шляхом зазначення на ній області інтересу. Стандарти метаданих для використання в ІПД розроблені як для міжнародного, так і для національного рівнів) [65].

Відповідно до прийнятих на державному та міжнародному рівні концепцій, ІПД являє собою ієрархічну систему, що має національний, регіональний і муніципальний рівні. Розглядають також можливості створення глобальної ІПД. Така структуру використовується при формуванні основної форми ІПД - національної. Національна ІПД (NSDI - англ.) дозволяє сформувати систему для доступу до базових просторових даних, що покриває територію однієї країни. Ця ІПД повною мірою вбудовується в систему державного територіального управління і відповідає його вертикальній і горизонтальній структурам [164].

1.4. Системний підхід та методологія обробки геопросторових даних

Управління територією - це управління системою високого ступеня складності (мегасистема). Для таких систем характерною ознакою є розгалужена структура і велике число зовнішніх і внутрішніх зв'язків. Особливістю мегасистем є відсутність можливості точного і докладного їх опису. Однак можна отримати загальне представлення про цю глобальну систему через інтерпретацію просторових даних, та її складових [21, 77, 80].

Простір складається з безлічі систем (підсистем), кожна з котрих міститься в більш масштабній системі. Внутрішня просторова політика держави (системи) реалізується шляхом впливу держави на соціально-економічний і політичний розвиток регіонів через підсистеми: територіальні органи державної влади (підрозділи міністерств, служб, наглядів, агентств, інспекцій); діяльність органів влади місцевого самоврядування трьох рівнів. Всі вони націлені на здійснення регіональної політики децентралізації.

В даний час сформувалися нові підходи щодо розуміння простору як геопростору. Геопростір - географічна оболонка Землі, що підлягає вивченню, відображенню, моделюванню територіальної локалізації в певні періоди часу, об'єктового складу, переліку та ступеня деталізації геоінформації, зазначених її споживачем. Простір мікросвіту і космічний простір не включаються в це поняття. Іншим критерієм, обмежуючим включення просторів в поняття «географічного», є застосування їх для вивчення та моделювання координатних систем: географічних і геодезичних.

Геопростір, як різновид простору, характеризується протяжністю, динамічністю, структурністю, безперервністю, конвертованістю.

Протяжність геопростору характеризується територіальним охопленням планети, півкулі, континентів і океанів, групи держав, країни, географічної провінції (наприклад, Західна Україна, Буковина), одиниці адміністративно-територіального поділу, населеного пункту, його частини або іншим.

Динамічність геопростору обумовлена його мінливістю і нерозривним зв'язком з часом. Ця обставина, з одного боку, дозволяє повідомляти як про

минулий стан аналізованого простору, так і про майбутнє (у вигляді проекту, прогнозу), з іншого боку, вимагає фіксації моменту вивчення геопростору, або його складових.

Структурність геопростору проявляється в наявності і розташуванні об'єктів геопростору (у тому числі, явищ і проявів процесів), що містяться і відбуваються в геопросторі в кожен конкретний момент часу. Прикладами процесів, що відбуваються в геопросторі є: повені, поширення захворювань, забруднення навколишнього середовища, військові операції, демографічні процеси, розподіл температури та ін. Об'єкти геопростору можуть бути конкретними (наприклад, населені пункти) і абстрактними (щільність населення), реальними (річкова мережа) і передбачуваними (проектowana мережа зрошення).

Безперервність геопростору характеризується наявністю в кожній його точці якого-небудь об'єкта геопростору; Не існує «порожнього» геопростору. Тим не менш, геопростір існує в декількох вимірах, в тому числі у фізичному і віртуальному.

Конвертованість геопростору - це можливість віртуальної трансформації з одного геопростору в інший. У цьому полягає одна з фундаментальних властивостей геопростору, як різновиду кіберпростору.

Методологічні підходи до інтелектуального аналізу геопросторових даних почали свій розвиток з середини XX сторіччя, коли разом з індустріальним розвитком економіки з'явилась необхідність у дослідженні оптимальних підходів до просторової організації суспільних процесів та охорони і раціонального використання природних ресурсів. Одночасно з цим з кінця 60-х років починається розвиток геоінформаційних систем як інформаційної основи для впровадження методів аналізу даних. Поява високопродуктивних обчислювальних машин разом з розвитком геопросторової індустрії та загостренням проблем економічного розвитку та екологічної ситуації призвели до активізації досліджень інтелектуального аналізу геопросторових даних та виділення його в окремий напрям [129].

В Україні таким поштовхом стала катастрофа на Чорнобильській АЕС, яка спонукала широке коло фахівців до використання геоінформаційних технологій при аналізі наслідків радіоактивного забруднення [73]. В галузі інтелектуального аналізу даних в Україні широко відомі роботи Інституту прикладного системного аналізу НТУУ «КПІ», Світового центру даних з геоінформатики та сталого розвитку Згуровського М.З., Панкратова Н.Д., А.І.Петренко, Болдака А.О. [71, 111]. Академічні та університетські дослідження інтелектуального аналізу відображені в роботах Зацерковного В.І. [64], Сарычевої Л.В. [168], Ситника, В.Ф., Краснюка М.Т., Черняка О.І., Захарченка П.В. та ін. [172,189].

В ХХІ столітті інтелектуальний аналіз геопросторових даних сформувався в окремий перспективний напрям наукових досліджень, який має свою структуру та особливості використання. Міжнародна дослідницька школа ІАГД сформувалась під впливом робіт таких вчених, як Andrienko, Cleve, C., Gahegan, Guo, Peuquet, & Gahegan, Guo, Han, Koperski & Stefanovic, Kulldorff, Mennis & Liu, Miller & Han, Openshaw, Charlton, Wymer, & Craft, Shekhar, Shiode, S. [196, 225, 269, 270, 272, 292, 296, 314, 318, 328, 372, 375] та ін.

Інтелектуальний аналіз геопросторових даних має глибоке коріння в традиційних областях геопросторового аналізу (наприклад, просторова статистика, аналітична картографія, пошуковий аналіз даних) і різних методах інтелектуального аналізу даних в статистиці та інформатиці. Його мета полягає в інтеграції і подальшому вдосконаленні методів в різних областях для аналізу великих і складних наборів геопросторових даних [129].

В якості визначення можна запропонувати наступне:

Інтелектуальний аналіз геопросторових даних (spatial data mining) – напрям інтелектуального аналізу даних, що направлено на системний аналіз геопросторових даних з метою пошуку закономірностей та відносин в даних, які засновані на просторовому розташуванні та топологічних відношеннях [129].

ІАГД є зростаючим дослідницьким напрямом, який все ще перебуває на ранніх стадіях розвитку та оперує цілим рядом методів, які комбінують обчислювальні, статистичні та візуальні методики. Інтелектуальний аналіз

геопросторових даних та виявлення знань є ітераційним процесом, що включає в себе кілька етапів, включаючи відбір даних, фільтрацію, первинну підготовку та перетворення; аналіз за допомогою обчислювальних алгоритмів і / або підходів з візуалізації, інтерпретації та оцінки результатів; формулювання або модифікацію гіпотез і теорій; корегування даних і методів аналізу; повторну оцінку результатів [129].

Інтелектуальний аналіз даних та виявлення знань має пошукову природу і є більш індуктивним процесом, ніж традиційні статистичні методи. Це природно вписується в початкову стадію дедуктивного процесу дослідження, де дослідники розробляють та модифікують теорії, що базуються на відкритій інформації за даними спостережень) [316-321].

ІАГД знаходиться на перетині ГІС, Data Mining, статистики та прикладних галузей досліджень. З цієї точки зору використання інструментів інтелектуального аналізу даних разом з геоінформаційним блоком є значною перевагою обробки вхідної інформації для управління територією та використання ГІС в СППР. У процесі застосування геоінформаційних технологій в СППР територіального управління можна виділити три базові рівні. До першого, інформаційного рівня відноситься вхідна інформація, яка використовується для обробки в ГІС. На сьогодні інтелектуальні системи можуть обробляти як структуровані, так і неструктуровані дані. Специфічними для використання в ГІС є бази геопросторових даних, дані дистанційного зондування Землі, дані із датчиків та сенсорів з відомими координатами, геотегована інформація із Інтернет та соціальних мереж, історичні картографічні дані, волонтерська географічна інформація та ін. (рис. 1.11) [129].

Поряд із ними широко використовуються механізми інтеграції просторової та непросторової інформації на основі процесів географічної прив'язки та геокодування. Це надає змогу повноцінно використовувати у ГІС інформацію із зовнішніх непросторових баз даних, які супроводжуються та наповнюються іншими органами управління. Необхідною складовою є бази знань структуровані за предметними галузями, які використовуються у СППР [324-331].

Аналітичний блок спирається на використання геоінформаційної системи як платформи для обробки та аналізу вхідної інформації. Тут можуть бути виділені функції стандартизації та нормалізації, сумісної обробки, інтелектуального аналізу даних. На цьому етапі ГІС може взаємодіяти з іншими системами обробки даних та бути складовою СППР [159].

Ситуаційний рівень пов'язаний з необхідністю прийняття рішень при аналізі конкретних ситуацій та сценарного моделювання за допомогою інструментів СППР та ГІС. Важливим компонентом в цьому випадку стають механізми колективної участі в роботі за участі керівних осіб різних рівнів, експертів технічного персоналу. До таких механізмів відносяться ситуаційні центри, он-лайн конференції, робочі наради та ін.

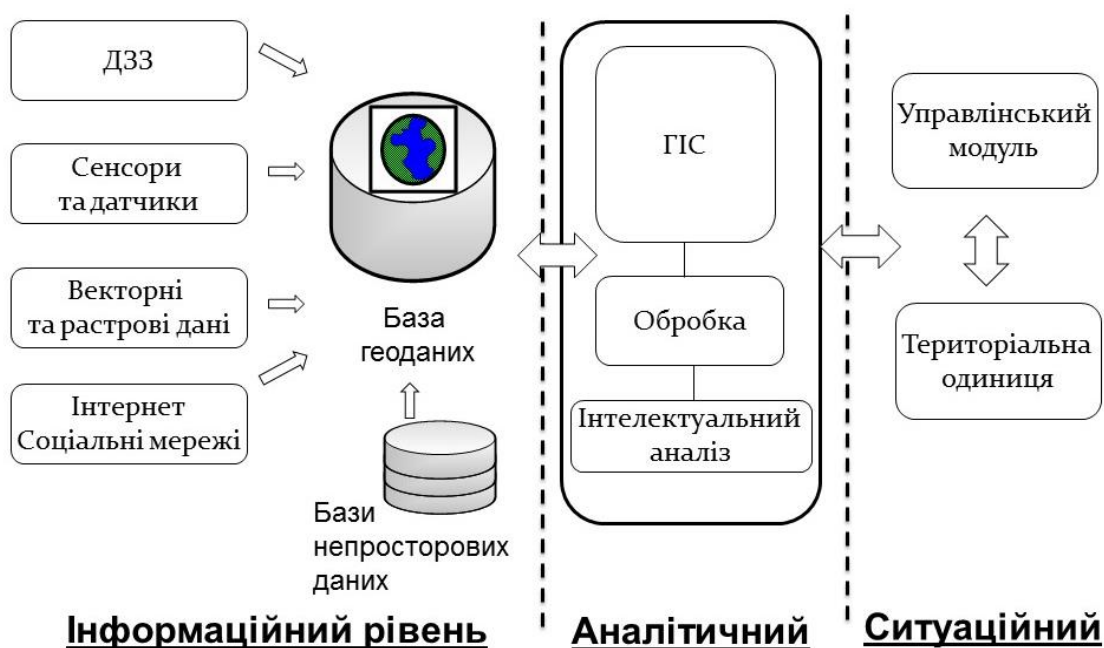


Рис. 1.11. ІАГД в структурі СППР територіального управління

Процес ІАГД містить основні кроки перетворення даних. Кожний аналіз здійснюється виконавцем з певною метою на основі вхідних баз даних та баз знань, які відповідають проблемному напрямку. На першому кроці обробка даних здійснюється за допомогою інтерфейсу СКБД. На цьому етапі здійснюється вибірка даних з усієї сукупності збереження даних, проведення простих операцій аналізу даних, які можливо здійснити на основі вбудованих засобів СКБД. На

другому кроці відбувається обробка даних за допомогою алгоритмів ІАГД або їх послідовне використання. Отримані результати аналізуються з метою виявлення нової інформації та знань [345]. За результатами цього аналізу відбувається оцінювання результатів обробки та даних і в залежності від того, чи відповідають отримані результати критеріям аналізу процедура аналізу повторюється з іншими параметрами, або отримані результати передаються для прийняття на їх основі рішень (рис.1.12).

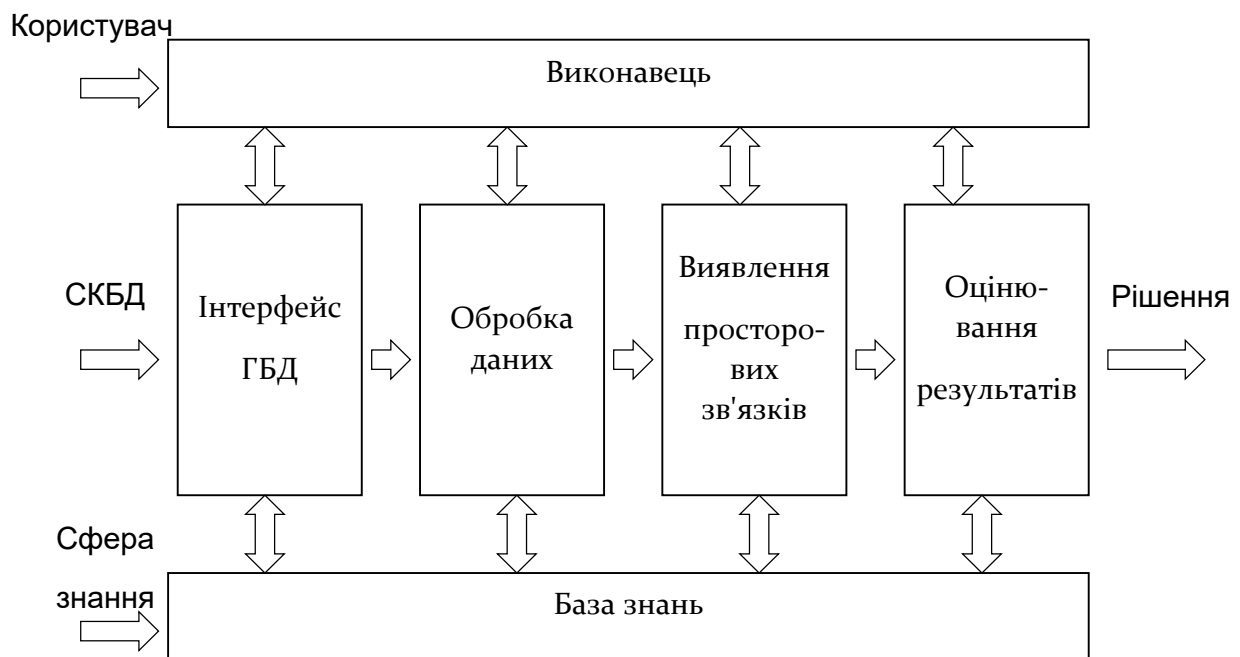


Рис.1.12. Процес обробки даних в ІАГД

До ІАГД відноситься декілька напрямів, які дозволяють провести різностороннє дослідження даних з урахуванням геопросторової складової. До їх числа відносяться методи класифікації геопросторових даних, кластеризація та регіоналізація, правила асоціацій, геостатистика, геовізуалізація. Цей перелік продовжує розширюватися разом із розвитком ІАГД. Кожний із цих напрямів містить сукупність методів, за допомогою яких відбувається аналіз даних (рис.1.13) [159].



Рис.1.13. Класифікаційна структура методів ІАГД.

Просторова класифікація та прогнозування

Класифікація є процесом групування елементів даних за класами (категоріями) відповідно до їх властивостей (значення атрибутів). Класифікацію також називають керованою класифікацією, на відміну від некерованої класифікації (кластеризація). Керована класифікація потребує початковий набір даних для навчання (або налаштування) моделі класифікації, набір даних перевірки для підтвердження (або оптимізації) конфігурації, і тестовий набір даних, щоб оцінити якість моделі) [365-368].

До методів класифікації належать, наприклад, дерева рішень, штучні нейронні мережі (ШНМ), метод максимальної правдоподібності (ММП), лінійний дискримінантний аналіз (ЛДА), метод опорних векторів (англ. SVM), методи найближчого сусідства і методи, що ґрунтуються на основі міркування (англ. CBR).

Просторові методи класифікації розширюють загальні методи класифікації враховуючи не тільки атрибути об'єкту, що класифікуються, але й атрибути сусідніх об'єктів та їх просторові відносини. Серед таких розробок слід звернути увагу на візуальний підхід до просторової класифікації, який було впроваджено в

дерево рішень засноване на традиційному алгоритмі C4.5, що поєднується з картою візуалізації. Це дозволяє виявити просторові закономірності правил класифікації. Індукція дерева рішень також використовується для аналізу і прогнозування просторового вибору поведінки [19]. Дистанційне зондування є одним з основних напрямків, в якому зазвичай використовують методи класифікації для розподілу пікселів за категоріями. Широке використання при аналізі пікселів здобули штучні нейронні мережі) [371-377, 392-394].

Правила асоціації у видобутку даних були спочатку призначена для виявлення закономірностей між записами у великих транзакціях баз даних. Подібно до пошуку правил асоціації в транзакційних або реляційних базах даних, правила просторових асоціацій можуть бути знайдені в базі геопросторових даних, розглядаючи просторові властивості і предикати [196].

Правило просторової асоціації виражається у вигляді $A \Rightarrow B [s\%, c\%]$, де A і B є набори просторових або непросторових предикатів, $s\%$ є підтримка правила і $c\%$ є впевненість в правилі.

Багато можливих просторових предикатів (наприклад, «близько до», «далеко від», «перетинання», «перекриття» і т.п.), можуть бути використані в правилах просторової асоціації. Але це потребує розгляду дуже великого обсягу обчислень різних просторових предикатів при виводі асоціативні правил з великих обсягів просторових даних. Ще одна потенційна проблема з пошуком правил просторової асоціації полягає в тому, що може бути сформована велика кількість правил, багато з яких є очевидними або загальновідомими. Тому необхідна наявність предметних знань, щоб відфільтрувати тривіальні правила і зосередитися тільки на нових і цікавих відкриттях) [312 - 315].

Інтелектуальний аналіз просторової спільності розташування порівняно подібний, але технічно дуже відрізняється від правил асоціації. Враховуючи набір геопросторових даних і їх місце розташування, патерн спільності розташування являє підмножини об'єктів з певною частотою розташованих разом, наприклад, деякі види птахів, як правило, проживають поряд з певним типом дерев. Виміри та

алгоритми для інтелектуального аналізу просторових спільностей розташування структур були запропоновані в ряді зарубіжних робіт [318, 372].

Методологічно з правилами асоціацій тісно пов'язаний аналіз топології, який дозволяє встановлювати правила поведінки об'єктів один відносно іншого, а також за допомогою функцій околу та оверлейних операцій виявляти подібні правила у геопросторових даних [129].

Кластеризація та регіоналізація

Кластерний аналіз широко використовується як метод аналізу даних, який організує набори елементів даних в групах (або кластерах), таким чином, що елементи в одній групі більше схожі між собою та істотно відрізняються від елементів в інших угрупованнях. Методи *кластеризації* можна поділити на дві групи: неієрархічна та ієрархічна кластеризація [159].

Для розгляду кластеризації геопросторових даних було проаналізовано три типи кластеризації, в тому числі геопросторової кластеризації (кластеризація просторових точок), регіоналізація (кластеризація з географічними обмеженнями примикання) і аналіз точкового розміщення (статистично вагомі скупчення точок «hot spot» за статистикою просторового сканування). Для першого типу просторової кластеризації схожість між точками даних або кластерами визначається за просторовими властивостями (наприклад, місце і відстань). Просторові методи кластеризації можуть бути неієрархічними або ієрархічними, заснованими на основі розподілу щільності, або на основі рівнорозподіленої сітки комірок) [356-362].

Регіоналізація є особливою формою кластеризації, яка прагне поєднати групи просторових об'єктів в просторово суміжні кластери (тобто регіони) при оптимізації цільової функції. Багато географічних додатків, таких як кліматичне зонування, ландшафтний аналіз, сегментація даних дистанційного зондування, часто вимагають, щоб кластери були географічно суміжними. Існуючі методи районування, засновані на концепції кластеризації, можна розділити на три групи:

1) багатовимірна (непросторова) кластеризація, яка потім потребує додаткової просторової обробки з метою упорядкування кластерів в регіони;

2) кластеризація з виміром просторово зваженого відхилення, яка вважає просторові властивості фактором формування кластерів;

3) обмежена примиканням кластеризації, що слугує для формування просторової безперервності в процесі кластеризації [270].

Аналіз точкового розміщення, який також відомий як аналіз «гарячі місця», фокусується на виявленні незвичайних концентрацій подій в просторі, наприклад, географічних кластерів хвороб, злочинності, або дорожньо-транспортних пригод. Загальною дослідницькою проблемою є визначення того, чи існує надлишок спостережуваних точкових подій (наприклад, випадків захворювання) для області (наприклад, в межах певної відстані до місця). Для пошуку подібних просторових кластерів використовується декілька статистичних методів, таких як метод Getis-Ord Gi [329], географічна машина аналізу (ГМА) [296] і сімейство просторово-часових методів статистики локального сканування. Все частіше статистика для виявлення просторових кластерів доступна для неевклідових просторів, зокрема мережових просторів) [257].

В якості тестової статистики в ГМА використовується кількість точок (наприклад, випадки епідемічних хвороб) в межах певної області (тобто, кругової області навколо центру комірки). Щоб визначити, чи має значущість кількість точок в області, використовується процедура Монте-Карло для генерації великої кількості (наприклад, 500) випадкових наборів даних, кожний з яких представляє реалізацію нульової гіпотези в тій же самій області. Значення статистики випробувань розраховують для кожного випадкового набору даних i , таким чином, розподіл значень тестової статистики при нульовій гіпотезі є похідним. Порівнюючи фактичне значення тестової статистики (тобто кількість точок) і отримані результати розподілу, отримуємо рівень значущості для тестової статистики в області. Потенційною проблемою з ГМА, є те, що важко встановити рівень для завдання мультитестування. Обчислювальне навантаження також є недоліком, але в більшій чи меншій мірі всім статистичним додаткам сканування буде потрібна значна обчислювальна потужність для пошуку та тестування локальних кластерів) [293-295].

Статистика локального сканування, розроблена Кульдорфом обчислює відношення правдоподібності для кожної локальної області. Для подолання проблеми множинного тестування, статистика локального сканування використовує максимальне відношення правдоподібності (яка є максимальним коефіцієнтом ймовірності серед усіх місць), як тестову статистику. Таким чином, спосіб статистики локального сканування визначає найімовірніший кластер, хоча множина вторинних кластерів також додається. Для цього алгоритм спочатку обчислює відношення правдоподібності для кожної із сукупності зон і знаходить максимум. Для виводу рівня значущості, генеруються копії наборів даних при нульовій гіпотезі, умови якої розповсюджуються на загальну кількість точок. Для кожної реплікації, значення тестової статистики обчислюють знову (тобто максимальне відношення правдоподібності знаходиться над усіма перерахованими місцях). Тоді фактичне значення тестової статистики порівнюється з тестовими значенням всіх реплікацій для отримання рівня значущості для найбільш ймовірного кластеру, а також вторинних кластерів) [378].

Характеризувати розподіл точок можна з використанням статистичних індексів. Прикладом такого підходу є діаграма Морішіта. Індекс Морішіта обчислюється для області, розбитою на рівнокутні комірки рівного розміру, за формулою [22]:

$$I_{\delta} = Q \frac{\sum_{i=1}^Q n_i(n_i-1)}{N(N-1)}, \quad (1.17)$$

де N - повне число точок мережі моніторингу; Q - кількість осередків розбиття; n_i ($i = 1, 2, \dots, Q$) - число точок мережі моніторингу, що потрапили в i -ту комірку. Цей індекс характеризує ймовірність того, що при виборі двох випадкових точок вони опиняться в одній комірці. Існують три типи характерних значень індексу Морішіта, комбінації яких дозволяють судити про характеристики мережі моніторингу:

- величина індексу Морішіта із зростанням розміру комірки зростає і прагне до 1, тоді розподіл точок можна вважати рівномірним;

- величина індексу Морішіта не залежить від розміру комірки і приблизно дорівнює ≈ 1 (коливається близько 1); це означає, що розподіл точок випадковий і не має кластерних структур.
- величина індексу Морішіта із зростанням розміру комірки зменшується або росте вище 1 – розподіл точок мережі кластерний [159].

Геовізуалізація

Цей напрям стосується розвитку теорії та методології полегшення конструювання знань на основі візуального дослідження та аналізу геопросторових даних та реалізації візуальних інструментів для подальшого пошуку знань, синтезу, зв'язків та використання [269]. Як галузь, що активно розвивається, геовізуалізація залучає інтереси різних наукових областей і розвивається разом з різноманітним набором напрямів досліджень. Основна відмінність між традиційною картографією та геовізуалізацією полягає в тому, що, картографія фокусується на розробці та використанні карт для передачі інформації та суспільного споживання, а геовізуалізація підкреслює розвиток інтерактивних карт і пов'язаних з ними інструментів для дослідження даних, побудови гіпотез і створення нових знань [159].

Геовізуалізація також має тісні зв'язки з дослідницьким аналізом даних (EDA) і пошуковим просторовим аналізом даних (ESDA), які пов'язують статистичні графіки і карти і спираються на вміння людини-експерта взаємодіяти з даними, візуально виявляти закономірності і формулювати гіпотези/моделі) [221]. Однак, щоб впоратися з сучасними великими і різноманітними наборами геопросторових даних і полегшити відкриття і розуміння складної інформації, геовізуалізації необхідно вирішити ряд завдань, у тому числі:

- 1) ефективна обробка дуже великих наборів даних (big data);
- 2) обробка декількох сценаріїв і робота з багатьма змінними одночасно, що дозволяє виявити складні зв'язки між даними;
- 3) розробка ефективних користувацьких інтерфейсів і інтерактивної стратегії полегшення пошукового процесу.

Для обробки великих масивів даних і візуалізації загальних закономірностей, візуальні підходи часто поєднуються з обчислювальними методами (наприклад, кластеризація, класифікація та правила асоціації) для узагальнення даних, виділення структур і допомоги користувачам при вивченні і розумінні закономірностей [196]. Для візуалізації різних точок зору і багатьох змінних часто необхідно об'єднувати методи візуалізації з методами зменшення розмірності, такими як багатовимірне шкалювання, аналіз головних компонент (РСА), карти самоорганізації, або інші методи пошуку оптимального відображення даних. Багатовимірне картографування є актуальною проблемою дослідження, для якого були розроблені численні підходи, такі як спеціально розроблені символи, мультизв'язані перегляди і підходи, які ґрунтуються на кластеризації [272]. Дослідницькі зусилля третьої хвилі з'явилися як активний субнапрямок під назвою візуальної аналітики) [299 - 302].

В межах дослідження було розроблено рішення шести прикладних задач моделювання складових сталого розвитку на основі моделей та методів обробки геопросторових даних (рис. 1.14).



Рис. 1.14. Структура задач, що вирішуються в межах дослідження

В роботі вирішено задачі наступних класів:

7. Визначення впливу просторових відносин на показники сталого розвитку
8. Відновлення пропусків значень у статистичних рядах метрики сталого розвитку
9. Виділення однорідних територіальних утворень для реалізації заходів регіональної політики
10. Аналіз безпекового виміру на різних територіальних рівнях
11. Аналіз екологічної компоненти землекористування
12. Розрахунок метрики сталого розвитку на основі даних ДЗЗ

Задача 1 вирішується моделюванням просторових відносин на основі континуальної моделі структурування даних. Запропоновано методи побудови трендових моделей для аналізу показників сталого розвитку з використанням алгоритмів побудови та ранжування трендових поверхонь 1-3 ступенів.

Задача 2 ґрунтується на застосуванні інтерполяційних моделей для визначення невідомих значень методом побудови інтерполяційних моделей з підбором варіаграми на основі порівняльного аналізу детермінованих та стохастичних методів інтерполяції.

Задача 3 моделюється здійсненням регіоналізації регіонів України у багатовимірному просторі на основі методів кластеризації з врахуванням просторових відносин алгоритмами нечіткої самоорганізації k-середніх.

Задача 4 вирішує проблему аналізу великих об'ємів геопросторових даних з часовою прив'язкою на основі групування великих просторово-часових даних методом організацій в просторово-часовому кубі спільно з кластеризацією гарячих точок та використанням непараметричного тесту Манна-Кендалла і статистики Getis-Ord Gi.

Вирішення задачі 5 побудовано на класифікації адміністративних одиниць за співвідношенням типів землекористування із застосуванням карт самоорганізації Кохонена на основі нейронної мережі з побудовою карт самоорганізації.

Задача 6 вирішується застосуванням нейронних мереж для аналізу економічної компоненти сталого розвитку методом навчання згорткових нейронних мереж з градієнтним алгоритмом навчання.

Інструментарій вирішення поставлених задач ґрунтується на використанні геоінформаційної системи ArcGIS з відповідними розробленими бібліотеками на основі мови програмування Python та бібліотеки просторового аналізу Arcpy.

Висновки до розділу

На основі концепції сталого розвитку та доробків у сфері математичного оцінювання процесу сталого розвитку на національному та регіональному рівні розроблено методологію інтелектуального аналізу геопросторових даних для цілей сталого розвитку. Встановлено структурні та функціональні зв'язки між математичною мірою процесу сталого розвитку та просторово-часовими характеристиками розподілу, що дає можливість визначити основні закономірності та приховані залежності у даних.

Визначено основні складові концепції сталого розвитку та метрики її оцінювання. Метрика оцінювання включає в себе розрахунок індексу сталого розвитку на основі векторних додатків компоненту якості та безпеки життя населення. Визначено систему зв'язків між територіальними утвореннями, метрикою оцінювання сталого розвитку та застосуванням прикладних методів ІАГД.

Дано визначення основним характеристиками, які визначають властивості геопросторових даних, в тому числі поняття про геопростір, геопросторові дані, систему географічних координат.

Виконано огляд і аналіз досліджень з питань структури організації, джерел надходження та специфіки застосування у системному аналізі геопросторових даних. Сформульовано системні засади розробки методології використання геопросторових даних для здійснення ІАГД в задачах сталого розвитку.

Визначено, що до основних джерел геопросторових даних відносяться векторні та растрові формати даних, які зберігаються у вигляді баз геоданих, дані дистанційного зондування Землі, матеріали польової інструментальної зйомки, геотегована інформація у мереж Інтернет.

В якості основної системи обробки геопросторових даних обрано геоінформаційну систему з функціями збору, зберігання, аналізу та візуалізації геопросторових даних, а також інтеграція з іншими інформаційними платформами для сумісної обробки даних. Інтегрованою основою для сумісної обробки геоданих виступає інфраструктура геопросторових даних (ІПД) - ієрархічна впорядкована система, побудована з використанням кібертехнологій та геоінформаційних підсистем, заснована на загальних просторових даних і метаданих, а також мережі географічних інформаційних вузлів - геопорталов і каталогів метаданих.

Сформульовано визначення інтелектуального аналізу геопросторових даних на основі підходу до сукупності методів інтелектуального аналізу з врахуванням особливостей обробки даних, які пов'язані з просторовим місцеположенням та топологічними відносинами між об'єктами.

Інтелектуальний аналіз геопросторових даних є перспективним напрямом наукових досліджень, який розвивається на перетині геоінформатики, методів Data Mining та статистичних методів обробки даних із широким залученням методології системного аналізу. Визначено, що основною метою використання ІАГД є пошук закономірностей та відносин у масивах даних, які містять просторово координовану прив'язку. Тому використання ІАГД в складі СППР в територіальному управлінні та прогнозуванні є важливим та актуальним інструментом обґрунтування управлінських рішень. З'ясовано, що процес реалізації ІАГД відрізняється стадійністю та ітераційністю з метою пошуку оптимальних результатів аналізу. Сформовано класифікацію напрямів та методів ІАГД, до основних з яких відносяться класифікація, кластеризація, правила асоціацій, геостатистика та геовізуалізація, які разом формують методологію системного інтелектуального аналізу з метою підтримки прийняття рішень. ІАГД підвищує ефективність обробки даних разом з іншими методами аналізу даних на основі різних інформаційних платформ.

Практичне застосування ІАГД запропоновано на основі вирішення ряду задач оцінювання складових сталого розвитку, серед яких виділено рішення шести задач на основі моделей, методів, та алгоритмів ІАГД, які дозволяють удосконалити та підвищити точність метрики сталого розвитку.

В якості інструментальної платформи для розробки інструментарію ІАГД обрано геоінформаційну платформу ArcGIS компанії ESRI та мову розширення функціональності прикладних додатків Python.

РОЗДІЛ 2. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ НА ОСНОВІ ГЕОПРОСТОРОВИХ ДАНИХ

2.1. Базові моделі представлення геоданих

Інформаційна основа ГІС формує цифрові представлення (моделі) реальності. Розглядаючи дані по відношенню до описаних ними об'єктів, можна казати про цифрові об'єкти моделей, а також при застосуванні простору об'єктів в ГІС - про цифрові моделі об'єктів простору. Цифрові за формою, за своєю суттю моделі простору відносяться до типу інформаційних моделей, відмінних від реальних (наприклад, фізичних), математичних або моделей особливого типу, наприклад картографічних.

Об'єктом інформаційного моделювання в ГІС є просторовий об'єкт. Це одна з ключових концепцій геоінформатики. Він може бути визначений як цифрове представлення (модель) об'єкта реальності (місцевості), що містить його місце розташування та набір властивостей (характеристики, атрибути), або сам цей об'єкт.

Певна множина цифрових даних про просторові об'єкти утворює просторові дані. Вони складаються з двох взаємозв'язаних частин: позиційної (топологічно-геометричної) та непозиційної (атрибутивної) складових, які формують опис простору та тематичного змісту даних.

Просторові об'єкти як абстрактні уявлення реальних об'єктів і предмет інформаційного моделювання (цифрового опису) в ГІС традиційно класифікуються відповідно до характеру просторової локалізації відображених ними об'єктів реальності, розмірності простору, який вони утворюють, моделі даних, які використовують для їх опису та за іншими ознаками. В рамках такої об'єктної моделі даних вони можуть об'єднуватися в нові класи об'єктів, що відмінні від базових або створених раніше. Базовими (елементарними) типами просторових об'єктів, якими оперують сучасні ГІС, зазвичай вважаються:

- *точка* (точковий об'єкт) – 0-вимірний об'єкт, що характеризується плановими координатами;

- *лінія* (лінійний об'єкт, полілінія) – 1-вимірний об'єкт, утворений послідовністю не менше двох точок з відомими плановими координатами (лінійними сегментами або дугами);
- *область* (полігон, полігональний об'єкт, контур, контурний об'єкт) – 2-вимірний (площинний) об'єкт, внутрішня область, обмежена замкнутою послідовністю ліній (дуг в векторних топологічних моделях даних або сегментів в моделі «спагетті») і ідентифікується внутрішньою точкою (міткою);
- *піксель* (піксель, ПЕЛ) – 2-вимірний об'єкт, елемент цифрового зображення, найменша з його складових, що отримується в результаті дискретизації зображення (розбиття на далі неподільні елементи растру), елемент дискретизації координатної площини в растровій моделі (даних) ГІС;
- *комірка* (регулярна комірка) – 2-вимірний об'єкт, елемент розбиття земної поверхні лініями регулярної сітки;
- *поверхня* – 2-вимірний об'єкт, який визначається не тільки плановими координатами, а й аплікатою Z , яка входить в число атрибутів об'єктів, що утворюють її; оболонка тіла;
- *тіло* - 3-вимірний (об'ємний) об'єкт, що описується трійкою координат, включаючи аплікату Z , і обмежений поверхнями [179].

Загальний цифровий опис просторового об'єкта включає:

- найменування;
- опис місця розташування (місцезнаходження, локалізації);
- набір властивостей;
- «відносини» з іншими об'єктами;
- просторова «поведінка».

Два останніх елементи опису просторового об'єкта факультативні.

Найменуванням об'єкта служить його географічне найменування (власне ім'я, якщо воно є), його умовний код і / або ідентифікатор, який присвоюється користувачем або призначається системою.

Залежно від типу об'єкта його місце розташування визначається набором координат, організованим певним чином в рамках моделі даних. Це геометрична частина опису даних, геометрія (метрика) для опису просторових об'єктів, що відрізняється від їх семантики (непозиційних властивостей).

Перелік властивостей відповідає атрибутам об'єкта, якісним і кількісним його характеристикам, які приписуються йому в цифровому вигляді користувачем, можуть бути отримані в ході обробки даних або генеруються системою автоматично (до останнього типу атрибутів належать, наприклад, значення площі і периметрів полігональних об'єктів). Існує розширене тлумачення поняття атрибута об'єкта. Останньому можуть бути поставлені у відповідність будь-які типи даних: текст, цифрове зображення, відео і аудіозапис, графіка (включаючи карту), що, по суті, реалізується на практиці в мультимедійних електронних атласах. Таким чином, під атрибутами розуміються саме змістовні, тематичні (непозиційні, непросторові) атрибути (властивості) об'єктів.

Під відносинами розуміють перш за все топологію об'єктів. До топологічних властивостей просторових об'єктів прийнято відносити його розмірність (вимірність, просторову розмірність), згідно з якою вище були виділені нуль, одно-, дво- тривимірні об'єкти; замкнутість, якщо мова йде про лінійні об'єкти; зв'язність; простота (відсутність самонакладання лінійних об'єктів та «островів» в полігоні); знаходження на межі, всередині або поза полігоном; ознака точкового об'єкта, яка вказує, чи є він кінцевим для певної лінії. Прикладами топологічних «відносин» об'єктів є їх властивості «перетинатися» (або «не перетинатися»), «стосуватися», «бути всередині», «утримувати», «збігатися» та ін. [34].

Топологія разом з геометрією утворює тополого-геометричну частину опису даних, їх позиційну частину.

Таким чином, в найзагальнішому вигляді в просторових даних слід розрізняти і виділяти три складові частини: топологічну, геометричну і атрибутивну – «геометрію», «топологію» і «атрибутику» цифрової моделі просторового об'єкта.

Чіткий поділ позиційних і непозиційних даних – історична традиція, що має певне технологічне коріння. Управління атрибутивною частиною даних зазвичай покладається на засоби систем управління базами даних (СКБД), вбудованих в програмні засоби ГІС або зовнішніх по відношенню до них. У найбільш яскравій формі воно реалізовано в розглянутій нижче векторній моделі даних, атрибути якої представлені таблицею, зберігаються і управляються СУБД, що підтримує реляційну модель даних, а їх позиційна частина, пов'язана з атрибутивною через ідентифікатори просторових об'єктів та управляється іншими засобами. Моделі просторових даних такого типу отримали широке поширення і найменування *геореляційних*. Сучасною альтернативою цієї моделі є інтегрований підхід, коли і атрибутивна, і тополого-геометрична частини даних зберігаються і управляються в єдиному середовищі СУБД, а також об'єктний і об'єктно-реляційний підходи (і однойменні типи моделей даних).

Об'єктно-орієнтований підхід до моделювання просторових об'єктів вимагає також поняття їх «просторової поведінки».

Способи організації цифрових описів просторових даних прийнято називати моделями даних за традицією, успадкованою з теоретичних узагальнень проектування систем управління базами даних. Вони називаються також цифровими представленнями або просто представленнями просторових даних.

На концептуальному рівні всю множину моделей просторових даних можна розділити на три типи: моделі дискретних об'єктів, моделі неперервних полів і моделі мереж.

Побудувати вичерпну класифікацію моделей просторових даних навряд чи можливо, оскільки особливості модельованої предметної області та специфічні вимоги до функціональності ГІС можуть потребувати розробки і використання досить спеціальних моделей даних [177]. Надалі будемо використовувати набір базових моделей (представлень) просторових даних, що використовуються для опису об'єктів розмірності не більше двох (планіметричних об'єктів): растрова модель; регулярно-коміркова (матрична) модель; квадротомічна модель (квадродерево, дерево квадрантів, Q-дерево, 4-дерево); векторна модель: векторна

топологічна (лінійно-вузлова) модель; векторна нетопологічна модель (модель «спагетті»).

Растрова модель даних. Модель даних, яка має аналогії в комп'ютерній графіці, де растр - прямокутні комірки, що розбивають зображення на складові однорідні (гомогенні) далі неподільні частини – пікселями (від англ. Pixel, скорочення від «picture element» - елемент зображення), кожному з яких поставлений у відповідність деякий код, який ідентифікує колір в тій чи іншій системі кольорів (колірній моделі). З множини значень логічних пікселів складається цифрове зображення. Растрова модель даних в ГІС передбачає розбиття простору (координатної площини) разом з її просторовими об'єктами на аналогічні пікселям дискретні елементи, що впорядковані у вигляді прямокутної матриці. Для цифрового опису (позиціонування) точкового об'єкта при цьому буде досить вказати його приналежність до того чи іншого елементу дискретизації, враховуючи, що його положення однозначно визначено номерами стовпця і рядка матриці (при необхідності координати пікселя, або його центроїда або будь-якого кута можуть бути обчислені). Пікселу присвоюється цифрове значення, що визначає ім'я або семантику (атрибут) об'єкта. Аналогічним чином описуються лінійні і полігональні об'єкти: кожен елемент матриці отримує значення, відповідне до належності чи неналежності до нього того чи іншого об'єкта (Рис. 2.1) [26].

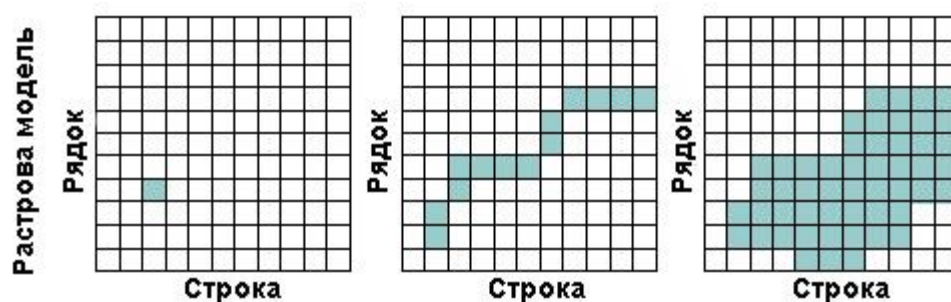


Рис. 2.1. Растрова модель даних

Обравши відповідний розмір пікселя растрової моделі, можна отримати просторове розрізнення (точність представлення об'єктів), що задовольняє цілям їх цифрового опису і подальшої обробки, якщо цьому не перешкоджають міркування економії машинної пам'яті: дворазове збільшення розрізнення веде до

чотириразового зростання обсягів збережених даних. Отримана матриця утворює растровий шар з однотипними об'єктами; множина різнотипних об'єктів утворює набір шарів, складових повного цифрового опису модельованої предметної області. З кожним семантичним значенням або кодом пікселя, крім того, може бути пов'язаний необмежений по довжині набір (таблиця) атрибутів, кожен з яких може бути розгорнуто як похідний шар з розмірністю вихідної матриці. Таким чином, стає не настільки обов'язковим поділ даних на позиційну і семантичну складові, відпадає необхідність в особливих засобах зберігання і маніпулювання метрикою і семантикою просторових даних, як це прийнято в векторних системах, істотно спрощуються аналітичні операції, багато з яких (включаючи обробку запитів з логічними умовами) зводяться до попиксельних операцій з набором растрових шарів, які можуть бути легко реалізовані паралельно.

У ГІС растрового типу (з варіантами підтримки растрової моделі даних) досить просто можуть бути реалізовані функції їх обробки, включаючи просторовий аналіз. Найчастіше вони включають також апарат, який отримав назву «картографічна алгебра», аналогічний за мовними засобами матричним операціям в деяких мовах програмування. Підтримка растрової моделі даних є передумовою інтеграції програмних продуктів ГІС із засобами цифрової обробки даних дистанційного зондування та обробки зображень в цілому [179].

Векторні моделі даних. Узагальнений клас векторних моделей включає два типи: векторні топологічні і векторні нетопологічні моделі. Вони використовуються для цифрового представлення точкових, лінійних і площинних (полігональних) об'єктів за аналогією з картографією, де розрізняються об'єкти з точковим, лінійним і площинним характером просторової локалізації, що визначає вибір графічних засобів їх картографічного відображення, і історично пов'язані з технологіями цифрування карт, планів і іншої графічної документації.

Множина точкових об'єктів, що утворюють шар однорідних даних (наприклад, множина об'єктів відповідних населених пунктів), може бути представлена у векторному форматі у вигляді неупорядкованої (необов'язково впорядкованої) послідовності записів (рядків таблиці), кожна з яких містить три

числа: унікальний ідентифікаційний номер об'єкта (ідентифікатор), значення координати X та Y у системі умовних планових прямокутних декартових координат.

Лінійний об'єкт (в загальному випадку крива) або межа полігонального об'єкта можуть бути представлені у вигляді послідовності точок (проміжних точок), тобто набором лінійних відрізків прямих (сегментів), що утворюють полілінію. Розташування точок, що утворюють полілінію, буде залежати від структури вихідної кривої. Їх прив'язка до характерних точок кривої при досить дрібному кроці цифрування дозволяє дати досить точне її наближення. Два з можливих варіанти її цифрового запису (формату) ілюструються нижче на рисунку 2.2.

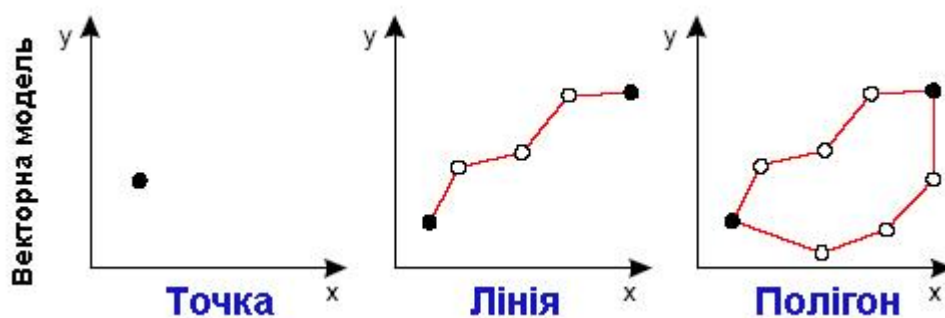


Рис. 2.2. Векторна нетопологічна модель

Таким же чином може бути представлена межа полігонального об'єкта. При цьому кожен іменований полігон (зі своїм ідентифікатором) представляється записом пар координат, що утворюють його межу в обраній послідовності. При описі множини полігонів кожен відрізок кордону, укладений між двома вузловими точками (за винятком зовнішнього кордону полігонів), буде описаний в цьому випадку двічі (за годинниковою стрілкою і проти). Така модель даних для опису точкових, лінійних і полігональних об'єктів носить найменування моделі «спагетті». Вона не є ефективною з точки зору ненадлишкових збережених даних і можливостей використання аналітичних операцій ГІС і підтримується недорогими програмними засобами настільного картографування і ГІС. Інше її найменування - *векторна нетопологічна модель*.

Векторна топологічна модель зобов'язана своїм походженням задачі опису полігональних об'єктів. Її називають ще лінійно-вузловий моделлю. З нею пов'язані особливі терміни, які відображають її структуру. Головні її елементи (примітиви):

- проміжна точка;
- сегмент (лінійний сегмент, відрізок (прямий));
- вузол;
- дуга;
- полігон (область, полігональний об'єкт, багатокутник,
- контур, контурний об'єкт), в тому числі:
 - простий полігон;
 - внутрішній полігон («острів», анклав);
 - складний полігон;
 - універсальний полігон (зовнішня область).

Опис полігону у векторній топологічній моделі – це множина трьох елементів: вузлів, дуг і полігонів. Між цими об'єктами встановлюються певні топологічні відносини, необхідним елементом яких повинен бути зв'язок дуг і вузлів, полігонів і дуг. Останнім приписуються покажчики розмежованих нею правого і лівого полігонів, конвенціалізуя напрямок обходу контурів (Рис. 2.3).

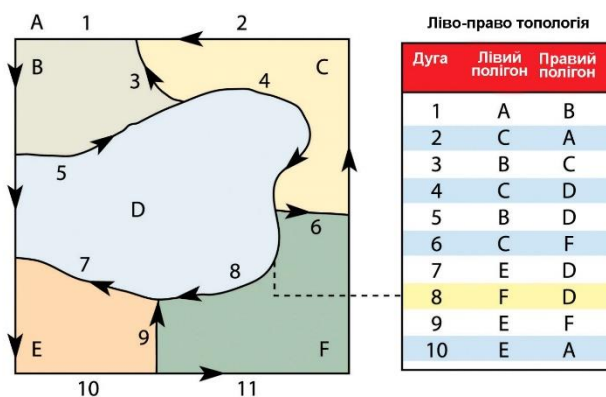


Рис. 2.3. Векторна топологічна модель

Багатомодельність програмного засобу, необхідність обміну даними між системами, інші мотиви спонукають розробку алгоритмів і засобів перетворення даних з однієї моделі в іншу. Деякі з цих перетворень прості і можуть бути виконані

автоматично, наприклад, векторно-растрове перетворення (растеризация). Зворотний процес – векторизація растрових даних (растрово-векторне перетворення), що широко застосоване при цифруванні графічних матеріалів, є набагато складнішим. Топологізація моделі «спагетті» – звичайна функція векторних редакторів, що підтримують перетворення нетопологічної векторної моделі в топологічну.

2.2. Базові методи інтелектуального аналізу геопросторових даних

До базових методів ІАГД відносяться методи, які ґрунтуються на визначенні залежностей між досліджуваними об'єктами на основі їх місцеположення та просторових відносин. Для параметризації таких типів взаємодії використовується нульова гіпотеза про розподіл об'єктів. Нульова гіпотеза для інструментів аналізу структурних закономірностей – це повна просторова хаотичність (ППХ) або самих об'єктів, або значень, пов'язаних з ними. Z-оцінки і р-значення, отримані в результаті аналізу структурних закономірностей, свідчать про можливість прийняття або відхилення нульової гіпотези. Як правило, для цього використовується один з інструментів аналізу структурних закономірностей, припускаючи, що z-оцінка і р-значення будуть свідчити про можливе спростування нульової гіпотези. Це буде говорити про те, що об'єкти або значення, пов'язані з ними, виявляють статистично значущу кластеризацію або дисперсію. У випадку, коли можна побачити просторову структуру, таку як кластеризація ландшафту (або просторових даних), можна бачити докази присутності деяких основних просторових процесів.

Р-значення – це імовірнісна характеристика. Для аналізу структурних закономірностей це імовірність того, що спостережувані просторові закономірності були створені певним випадковим процесом. Коли р-значення є дуже низьким, це означає, що існує дуже низька імовірність, що спостережувані

просторові закономірності є результатом випадкових процесів. Таким чином, можна відхилити нульову гіпотезу.

Z-оцінки є стандартним відхиленням у розподілі даних. Якщо, наприклад, інструмент повертає z-оцінку $+2.5$ – це 2.5 стандартних відхилення. z-оцінки, і p-значення пов'язані зі стандартним нормальним розподілом, як показано нижче.

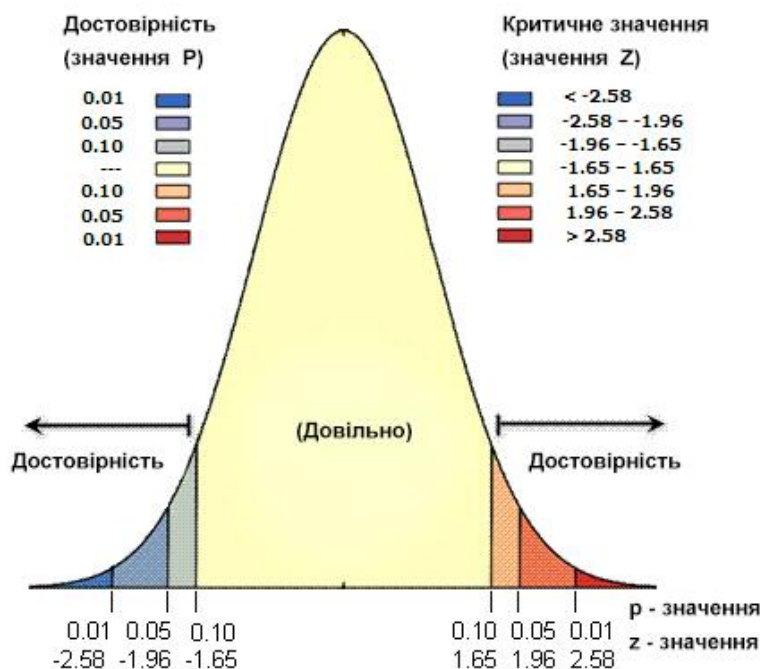


Рис. 2.4. Зразок нормального розподілу даних

Дуже високі або дуже низькі (негативні) z-оцінки, пов'язані з дуже низькими p-значеннями, розташовуються в хвостах нормального розподілу. Отримання малих p-значень з екстремально високими або низькими z-оцінками вказує, що спостережувана просторова модель відображає теоретичну випадкову структурну закономірність, яка представлена нульовою гіпотезою.

Щоб відхилити нульову гіпотезу потрібно зробити суб'єктивне судження щодо рівня ризику помилки (для того, щоб відхилити нульову гіпотезу). Перед запуском просторового статистичного процесу обирається довірчий рівень. Типові довірчі рівні 90, 95, або 99 відсотків.

У таблиці 2.1. нижче показані нескореговані критичні p-значення і z-оцінки для довірчих рівнів.

Таблиця 2.1.

Нескореговані критичні р-значення і z-оцінки для довірчих рівнів

z-оцінка (Стандартні відхилення)	р-значення (Імовірність)	Довірчий рівень
< -1,65 або > +1,65	< 0,10	90%
< -1,96 або > +1,96	< 0,05	95%
< -2,58 або > +2,58	< 0,01	99%

Розглянемо приклад. Критичні значення z-оцінки, використовуючи 95 % довірчий рівень є -1,96 і +1,96 стандартними відхиленнями. Нескореговане р-значення, яке пов'язане з 95 % довірчим рівнем, так само 0,05. Якщо z-оцінка знаходиться між -1,96 і +1,96, то нескореговане р-значення буде більше ніж 0,05 і ми не зможемо відхилити нульову гіпотезу, оскільки показана модель може, імовірно, бути результатом випадкових просторових процесів. Якщо z-оцінка потрапляє поза цим діапазоном (наприклад, -2,5 або +5,4 стандартних відхилень), просторова модель, яка спостерігається, імовірно, більш структурована, щоб бути результатом випадкового процесу, і р-значення будуть низькими. У цьому випадку можливо відхилити нульову гіпотезу і відновити пошук, що може викликати статистично істотну просторову структуру в даних.

Основна ідея полягає в тому, що значення в середині нормального розподілу (такі z-оцінки як 0,19 або -1,2), представляють очікуваний результат. Коли абсолютне значення z-оцінки є великим та імовірності є малими (в хвостах нормального розподілу), ми маємо справу зі статистично значущими «незвичайними» викидами у даних. Для алгоритму аналізу гарячих точок, наприклад, «незвичайний» означає статистично істотну «гарячу» або «холодну» точку.

Корекція FDR. Інструменти аналізу локальних просторових закономірностей, включаючи аналіз гарячих точок і локальний індекс Морана пропонують додатковий параметр застосування корекції FDR (Apply False Discovery Rate (FDR) Correction). Коли цей параметр включений Корекція FDR

знижує критичний поріг p -значення, показаний в таблиці вище, щоб використовувати в тестуванні множині просторової залежності. Якщо відбувається зменшення, то воно є функцією числа входних об'єктів і структури оточення, що використовується.

Алгоритми аналізу локальних просторових закономірностей працюють, розглядаючи кожен об'єкт в контексті навколишніх об'єктів, і визначають, чи відрізняється локальна закономірність (цільовий об'єкт і його оточення) від глобальної (всі об'єкти набору даних). Результати обчислень z -оцінки і p -значення, які пов'язані з кожним об'єктом, дозволяють визначити, чи є відмінність статистично значущою або ні. Цей аналітичний підхід створює певні складнощі при множинному тестуванні і вивченні залежностей. Актуальна проблема визначення закономірностей геопросторового розподілу індикаторів сталого розвитку та вирішення її в умовах, що формуються та неперервно змінюються.

У випадку множинного тестування з рівнем достовірності 95% теорія імовірності вказує, що існує тільки 5% зі 100%, що просторова закономірність може бути структурованою (кластеризованою або дисперсійною) і може бути пов'язана зі статистично значущим p -значенням, коли в реальності просторові процеси, що створюють цю закономірність, є випадковими. В цьому випадку ми невірно відкидаємо нульову гіпотезу, ґрунтуючись на статистично значущих p -значеннях. Наприклад, якщо є 10000 об'єктів, ми може отримати до 500 помилкових результатів.

Просторова залежність – близько розташовані об'єкти мають тенденцію до схожості; вони частіше, ніж не просторові дані, демонструють такий тип залежності. Проте, для багатьох статистичних тестів необхідно, щоб об'єкти були незалежними. Це необхідно для інструментів аналізу локальних закономірностей тому, що просторова залежність може штучно згладжувати статистичну значущість. Просторова залежність посилюється інструментами локального аналізу закономірностей, оскільки кожен об'єкт оцінюється в контексті його сусідства, і близько розташовані об'єкти матимуть множину однакових сусідств.

Для обробки проблем, що виникають з множинним тестом і просторовими залежностями, використовуються три підходи. Перший підхід - це ігнорувати проблему, з огляду на те, що окремий тест, виконаний для кожного об'єкта набору даних, повинен розглядатися окремо від інших. Однак при цьому підході досить імовірно, що деякі статистично значущі результати будуть невірні (виглядати статистично значущими при випадковому характері базових просторових процесів). Другий підхід полягає в застосуванні класичної процедури множинного тестування, наприклад поправки Бонферроні або корекції Сідак. Однак ці методи зазвичай дуже консервативні. Хоча вони значно знижують кількість помилково позитивних результатів, вони також пропускають наявні статистично значущі результати. Третій підхід полягає в застосуванні корекції FDR, яка оцінює число помилково позитивних результатів для даного рівня достовірності та відповідно коригує критичне р-значення. При цьому способі статистично значущі р-значення ранжуються від найменших до найбільших, на основі оцінки помилково позитивних результатів. Не будучи ідеальним, цей метод, як показують емпіричні тести, показує кращі результати, ніж виконання кожного тесту окремо або застосування традиційних методів множинного тесту.

Просторова автокореляція (Глобальний індекс І Морана) (Spatial Autocorrelation (Global Moran's I)) вимірює просторову автокореляцію, засновану одночасно на розташуванні об'єктів та їх значеннях. Виходячи із запропонованого набору об'єктів і пов'язаних з ними атрибутів, інструмент просторова автокореляція оцінює наявність кластеризації об'єктів або їх випадкового розподілу. Метод розраховує індекс І Морана, а також z-оцінку і р-значення, щоб оцінити значимість індексу.

Індекс Морана для просторової автокореляції задається як:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad (2.1)$$

де z_i – відхилення атрибута для об'єкту i від його середнього значення ($x_i - \bar{X}$), $w_{i,j}$ – просторова вага між об'єктами i та j , n – загальне число об'єктів і S_0 – сукупність всіх просторових ваг:

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j}, \quad (2.2)$$

z_I – оцінка для статистики розраховується як:

$$z_I = \frac{I - E[I]}{\sqrt{V[I]}}, \quad (2.3)$$

де:

$$E[I] = -1/(n - 1), \quad (2.4)$$

$$V[I] = E[I^2] - E[I]^2. \quad (2.5)$$

При розрахунку індексу розраховується середнє значення і дисперсія для атрибутів, які оцінюються. Потім зі значення кожного об'єкта віднімається середнє, створюючи відхилення від середнього. Значення відхилень для всіх сусідніх об'єктів (наприклад, об'єкти в межах зазначеного діапазону відстаней) перемножуються один на одний, щоб створити векторний добуток. Чисельник Глобального Індексу Морана включає суму добутоків векторів. Припустимо, що об'єкти А і В - сусіди, і середнє значення для них - 10. Зауважемо, що діапазон можливих добутоків векторів наступний:

Таблиця 2.2.

Добутки векторів Глобального Індексу Морана

Значення об'єктів		Відхилення		Векторні добутки
A=50	B=40	40	30	1200
A= 8	B=6	-2	-4	8
A=20	B=2	10	-8	-80

Коли значення сусідніх об'єктів вище середнього або нижче - векторний добуток буде позитивним. Коли одне значення менше середнього, а інше - більше, векторний добуток буде негативним. У всіх випадках, чим більше відхилення від середнього, тим більше векторний добуток. Якщо значення в наборах даних просторово кластеризуються (високі значення розташовуються поруч з іншими високими значеннями; низькі значення - поруч з іншими низькими); індекс Морана буде позитивним. Коли високі значення відкидають інші високі значення і розташовуються поблизу низьких значень, індекс буде негативним. Якщо позитивний векторний добуток балансує з негативним векторних добуток, індекс

буде близько 0. Чисельник нормалізується по варіації, тому Індекс приймає значення від -1 до +1.

Просторова автокореляція (Глобальний індекс І Морана) - це статистичний показник, який означає, що результати аналізу інтерпретуються в контексті нульової гіпотези. Для показника Глобального індексу Морана нульова гіпотеза стверджує, що аналізовані атрибути розподілені випадково між об'єктами або просторові процеси, які створили спостережувану структуру значень, носять випадковий характер.

Коли р-значення, отримане в результаті роботи даного інструменту, статистично низьке і статистично значуще, нульова гіпотеза може бути відхилена.

Таблиця 2.3. нижче представляє інтерпретацію результатів:

Таблиця 2.3.

Варіація Р та Z оцінок у розрахунку просторової кореляції

Р-значення статистично не значуще	Нульову гіпотезу не можна відхилити. Цілком можливо, що просторовий розподіл значень атрибутів об'єктів - результат випадкових просторових процесів. Спостережувана просторова структура значень може бути одним з можливих варіантів повної просторової хаотичності.
Р-значення статистично значуще і z-оцінка – позитивна	Нульову гіпотезу можна відхилити. Просторовий розподіл високих / низьких значень в наборі даних просторово більш кластеризований, ніж ми могли очікувати, якби зазначені просторові процеси були дійсно випадковими.
Р-значення статистично значуще, і z-оцінка - негативна.	Нульову гіпотезу можна відхилити. Просторовий розподіл високих і низьких

	<p>значень в наборі даних більш просторово розподілених, ніж це спостерігалось б, якщо зазначені просторові процеси були б дійсно випадковими. Дисперсна просторова модель часто відображає певний тип конкурентного процесу: об'єкт з високими значеннями розташовується поруч з іншими об'єктами з високими значеннями, об'єкт з низькими значеннями розташовується поруч з іншими об'єктами з низькими значеннями.</p>
--	---

Метрика середнього найближчого сусідства вимірює відстань між центроїдом кожного об'єкта і місцем розташування центроїда його найближчого сусіда. Потім вона усереднює ці відстані до найближчих сусідів. Якщо середня відстань менше середнього для гіпотетичного випадкового розподілу, вважається, що такий розподіл об'єктів кластеризується. Якщо середня відстань більше середнього для гіпотетичного випадкового розподілу, вважається, що такий розподіл об'єктів дисперсний. Співвідношення середнього найближчого сусідства розраховується як відношення спостережуваної середньої відстані до очікуваної середньої відстані (очікувана середня відстань розраховується для гіпотетичного випадкового розподілу з тією ж кількістю об'єктів, що покривають таку ж саму загальну область).

Середнє найближче сусідство розраховується за формулою:

$$ANN = \frac{\bar{D}_O}{\bar{D}_E}, \quad (2.6)$$

де \bar{D}_O – середня спостережувана відстань від кожного об'єкта до його найближчого сусіда.

$$\bar{D}_O = \frac{\sum_{i=1}^n d_i}{n}, \quad (2.7)$$

де \bar{D}_E – очікувана середня відстань від кожного об'єкта до його найближчого сусіда, яка розраховується наступним чином:

$$\bar{D}_E = \frac{0,5}{\sqrt{n/A}}. \quad (2.8)$$

В наведеному виразі d_i дорівнює відстані між об'єктом i та його найближчим сусідом, n відповідає загальній кількості об'єктів, та A є площа мінімального прямокутника, який охоплює всі об'єкти.

Z-оцінка середнього ближчого сусідства для статистики обраховується як:

$$Z = \frac{\bar{D}_O - \bar{D}_E}{SE}, \quad (2.9)$$

де SE (середньоквадратична похибка дорівнює):

$$SE = \frac{0.26136}{\sqrt{n^2/A}}. \quad (2.10)$$

Якщо індекс Z (співвідношення середнього найближчого сусідства) менше 1, тоді спостерігається кластеризація. Якщо Індекс більше 1, тоді спостерігається тенденція до дисперсності.

Рівняння, що застосовуються для розрахунку індексу Середнє найближче сусідство (2.6) і z-оцінка (2.9), засновані на припущенні, що вимірювані точки можуть розташовуватися в будь-якому місці в межах області дослідження (наприклад, відсутні межі і всі випадки або об'єкти розташовані незалежно один від одного). Р-значення - це числовий вираз області, розташованої під кривою, для відомого розподілу, обмеженого статистичним тестом.

Можливості застосування:

- Оцінка спірності території або інших результатів. Інструмент кількісно оцінює або порівнює просторовий розподіл різноманітності об'єктів в межах фіксованої області вивчення, порівнює середні найближчі відстані до сусідніх об'єктів для різних класів.

- Моніторинг змін в часі. Інструмент оцінює зміни в просторовій кластеризації для одного типу об'єктів в межах фіксованої області вивчення в часі.
- Порівняння спостережуваного і контрольного розподілів. Наприклад, при аналізі розподілу може знадобитися порівняти розміщення об'єктів для визначення того, чи відповідає їхня кластеризація очікуванням.

Аналіз кластерів і викидів визначає просторові кластери об'єктів з високими або низькими значеннями. Цей інструмент також визначає просторові викиди. Для цього обчислюється Локальний індекс Морана I , z -оцінка, p -значення, і код, який представляє тип кластера для кожного статистично значимого об'єкта. Z -оцінки і p -значення представляють статистичну значущість обчислених значень індексу.

Локальний індекс Морана статистично оцінює просторовий зв'язок між об'єктами:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X}), \quad (2.11)$$

де x_i – числовий атрибут об'єкту i \bar{X} – середнє значення для цього атрибута, $w_{i,j}$ – просторова вага для пари i та j , та:

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n (x_j - \bar{X})^2}{n-1} - \bar{X}^2, \quad (2.12)$$

де n відповідають загальній кількості об'єктів.

Оцінка z_{I_i} вираховується по наступній формулі:

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}}, \quad (2.13)$$

де:

$$E[I_i] = - \frac{\sum_{j=1, j \neq i}^n w_{i,j}}{n-1}, \quad (2.14)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2. \quad (2.15)$$

Додаткові розрахунки, надані нижче:

$$E[I^2] = A - B, \quad (2.17)$$

$$A = \frac{(n - b_{2_i}) \sum_{j=1, j \neq i}^n w_{i,j}^2}{n-1}, \quad (2.18)$$

$$B = \frac{(2b_{2_i} - n) \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n w_{i,k} w_{i,h}}{(n-1)(n-2)}, \quad (2.19)$$

$$b_{2i} = \frac{\sum_{i=1, i \neq j}^n (x_i - \bar{X})^4}{(\sum_{i=1, i \neq j}^n (x_i - \bar{X})^2)^2}. \quad (2.20)$$

Позитивне значення індексу вказує, що об'єкт має сусідні об'єкти з такими ж високими або низькими значеннями атрибута – тобто цей об'єкт є частиною кластера. Негативне значення індексу вказує, що об'єкт має сусідні об'єкти з різнорідними значеннями – тобто цей об'єкт є статистичним викидом. У будь-якому випадку, р-значення для об'єкта має бути досить малим, щоб кластер або викид можна було вважати статистично значущими. Зауважимо, що локальний індекс Морана – відносний вимір і може тільки інтерпретуватися в контексті обчислених для нього z-оцінки або р-значення.

Тип кластерів / викидів розрізняється на статистично істотний (0,05 рівня) кластер високих значень (HH), кластер низьких значень (LL), викид, в якому високе значення оточене в основному низькими значеннями (HL), і викид, в якому низьке значення оточене в основному високими значеннями (LH). Статистична значущість задається з рівнем достовірності 95%. Коли застосовується FDR корекція, об'єкти з р-значеннями, меншими 0,05 розглядаються як статистично значущі. Корекція FDR зменшує поріг р-значення з 0,05 до значення, яке краще відображає 95% рівень достовірності даного множинного тестування.

2.3. Методи просторового аналізу гарячих точок

Аналіз гарячих точок розраховує статистичний показник Getis-Ord G_i^* для кожного об'єкта в наборі даних. Підсумкові z-оцінки і р-значення говорять про те, в якій області простору кластеризуються об'єкти з високими або низькими значеннями. Інструмент працює шляхом аналізу кожного об'єкта в контексті сусідніх об'єктів. Об'єкт з високим значенням цікавий, але, можливо, не є статистично істотною гарячою точкою. Щоб бути статистично суттєвою гарячою точкою об'єкт повинен мати високе значення і бути оточений іншими об'єктами з також високими значеннями. Локальна сума для об'єкта і його сусідів порівнюється

пропорційно до суми всіх об'єктів. Коли локальна сума значно відрізняється від очікуваної локальної суми, і коли ця відмінність є занадто великою, щоб бути результатом випадкового процесу, виходить статистично значуща z-оцінка. Коли застосовується корекція FDR, статистична значимість налаштовується для множинного тестування і просторових залежностей

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{j,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}}, \quad (2.21)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}, \quad (2.22)$$

де x_j — атрибутивне значення для подій j , $w_{i,j}$ — просторова вага між подіями i та j , n — загальна кількість подій, \bar{X} — середнє арифметичне значень подій, S — дисперсія.

Статистична величина G_i^* , що присвоєна кожному об'єкту в наборі даних, є z-оцінкою. Для статистично значущих позитивних z-оцінок, чим більше z-оцінка, тим інтенсивніша кластеризація високих значень (гаряча точка). Для статистично значущих негативних z-оцінок, чим менше z-оцінка, тим інтенсивніша кластеризація низьких значень (холодна точка) [222].

Інструмент аналізу гарячої точки оцінює, чи кластеризуються в просторі високі або низькі значення (наприклад, число злочинів, локальних конфліктів). Поле, що містить ці значення, є полем аналізу. Проте, для точкових даних про інциденти, можливо більш цікавою буде оцінка інтенсивності інциденту, ніж аналіз просторового об'єднання в кластери будь-якого значення, пов'язаного з інцидентами. У цьому випадку необхідно агрегувати свої дані про інцидент до аналізу. Існує кілька методів виконання цього:

1. Обробку полігональних даних можна виконати з використанням операції Просторове з'єднання (Spatial Join), щоб підрахувати кількість подій в кожному полігоні. Отримане поле, що містить кількість подій в кожному полігоні, стає вхідним полем для аналізу.
2. Використання операцій Побудувати сітку для створення сітки полігонів поверх точкових об'єктів. Наступним етапом є використання операцій

Просторове з'єднання, щоб порахувати кількість подій, що потрапляють в межі кожного полігону сітки.

3. Якщо існує багато співпадаючих точок або пунктів на невеликій відстані один від одного, можна використовувати операції Інтегрувати і Зібрати події, щоб прив'язати об'єкти в межах вказаної відстані один від одного разом і потім створити новий клас об'єктів, що містить точку на кожному унікальному місці із пов'язаним числовим атрибутом, щоб ідентифікувати кількість подій / пов'язаних точок (Рис.2.5).

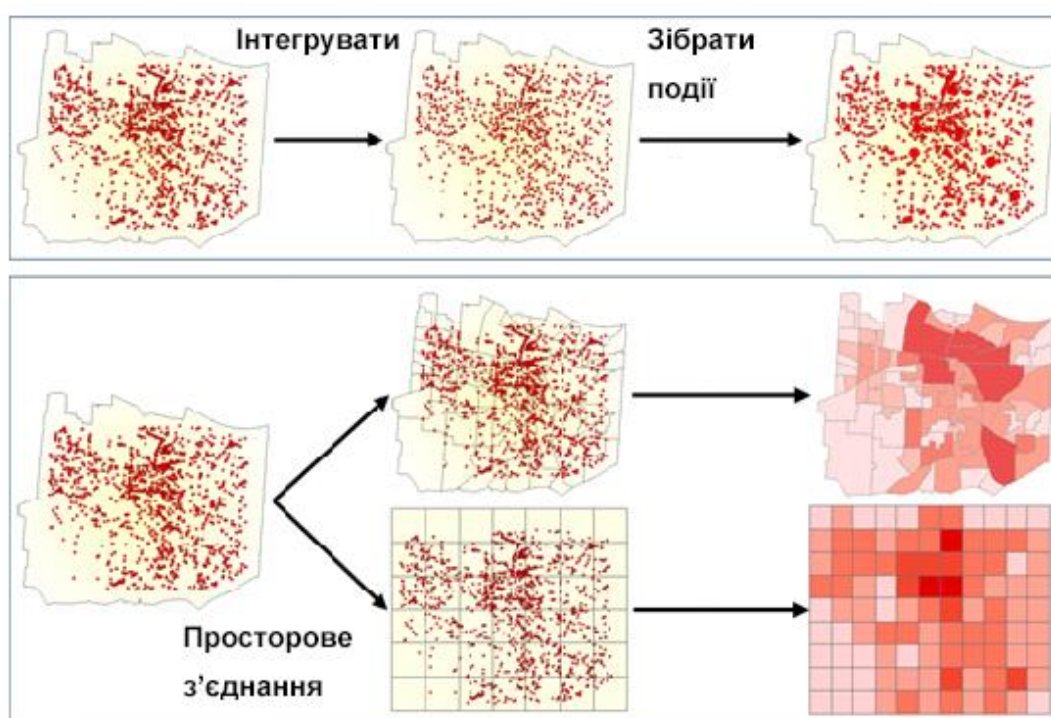


Рис. 2.5. Групування даних за просторовими ознаками

Аналіз групування (Grouping Analysis)

Базовими операціями аналізу даних є організація, групування, диференціація та класифікація, що дозволяє краще зрозуміти оточуючі об'єкти. Аналогічно для кращого вивчення даних можна використовувати інструментарій аналізу групування, який заснований на виконанні процедури класифікації, яка намагається знайти природні кластери в даних. Використовуючи вказану кількість груп, алгоритм шукає рішення, в якому всі об'єкти в кожній групі найбільш схожі, а самі групи максимально відрізняються одна від одної. Подібність об'єктів

заснована на наборі атрибутів, що вказуються для параметра аналізу, які при необхідності можуть накладати просторові або просторово-часові обмеження. При вказівці просторових або просторово-часових обмежень алгоритм використовує граф зв'язності (мінімальне дерево) для пошуку природних груп.

Хоча існує багато алгоритмів кластерного аналізу, всі з них класифікуються як NP-складні. При вирішенні не тільки не можна знайти оптимальне рішення проблеми, але також не можна визначити алгоритм угруповання, який буде працювати краще за всі для всіх можливих сценаріїв. Групи бувають різних форм, розмірів і щільності. Атрибути можуть містити дані з різними діапазонами, симетрією, неперервністю і одиницями виміру.

Ефективність угруповання об'єктів вимірюється за допомогою псевдо-F-статистики Калінскі-Харабаза, яка також відображає подібність об'єктів в групі і відмінність між групами:

$$\frac{\left(\frac{R^2}{n_c-1}\right)}{\left(\frac{1-R^2}{n-n_c}\right)}, \quad (2.23)$$

$$\text{де } R^2 = \frac{SST-SSE}{SST}, \quad (2.24)$$

і SST відображає різницю між групами, а SSE – схожість в середині групи:

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (V_{ij}^k - \overline{V^k})^2, \quad (2.25)$$

$$SSE = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (V_{ij}^k - \overline{V_i^k})^2, \quad (2.26)$$

де n – кількість просторових об'єктів

n_i – кількість просторових об'єктів у групі i

n_c – кількість класів (груп)

n_v – кількість змінних, що використовуються для групування об'єктів

V_{ij}^k – значення k -ої змінної для j -го просторового об'єкту в i -ій групі

$\overline{V^k}$ – середнє значення k -ої змінної

$\overline{V_i^k}$ – середнє значення k -ої змінної в групі i

Просторові обмеження

Щоб отримані групи були просторово близькі, задаються просторові обмеження, які відповідають підходам до концептуалізації об'єктів. Опції можуть визначати способи концептуалізації:

- спільне ребро;
- загальне ребро;
- вершина з іншим учасником групи.

Опції неперервності полігонів у випадку, якщо набір даних містить кластери несуміжних полігонів або полігонів без суміжних сусідів (рис. 2.6):



Рис. 2.6. Кластеризація полігональних об'єктів із ознаками примикання

Тріангуляції Делоне та K -найближчий сусід вказують, що об'єкт включається в групу, тільки якщо хоча б один інший об'єкт є природним сусідом (тріангуляція Делоне) або K найближчим сусідом.

Тріангуляції Делоне не використовується для наборів даних з однаковими об'єктами. Крім того, так як метод тріангуляції Делоне перетворює об'єкти в полігони Тіссена для визначення просторових відносин, особливо коли в наборі даних є полігональні об'єкти та іноді периферійні об'єкти, результати використання цієї опції не завжди будуть очікуваними (Рис. 2.7).

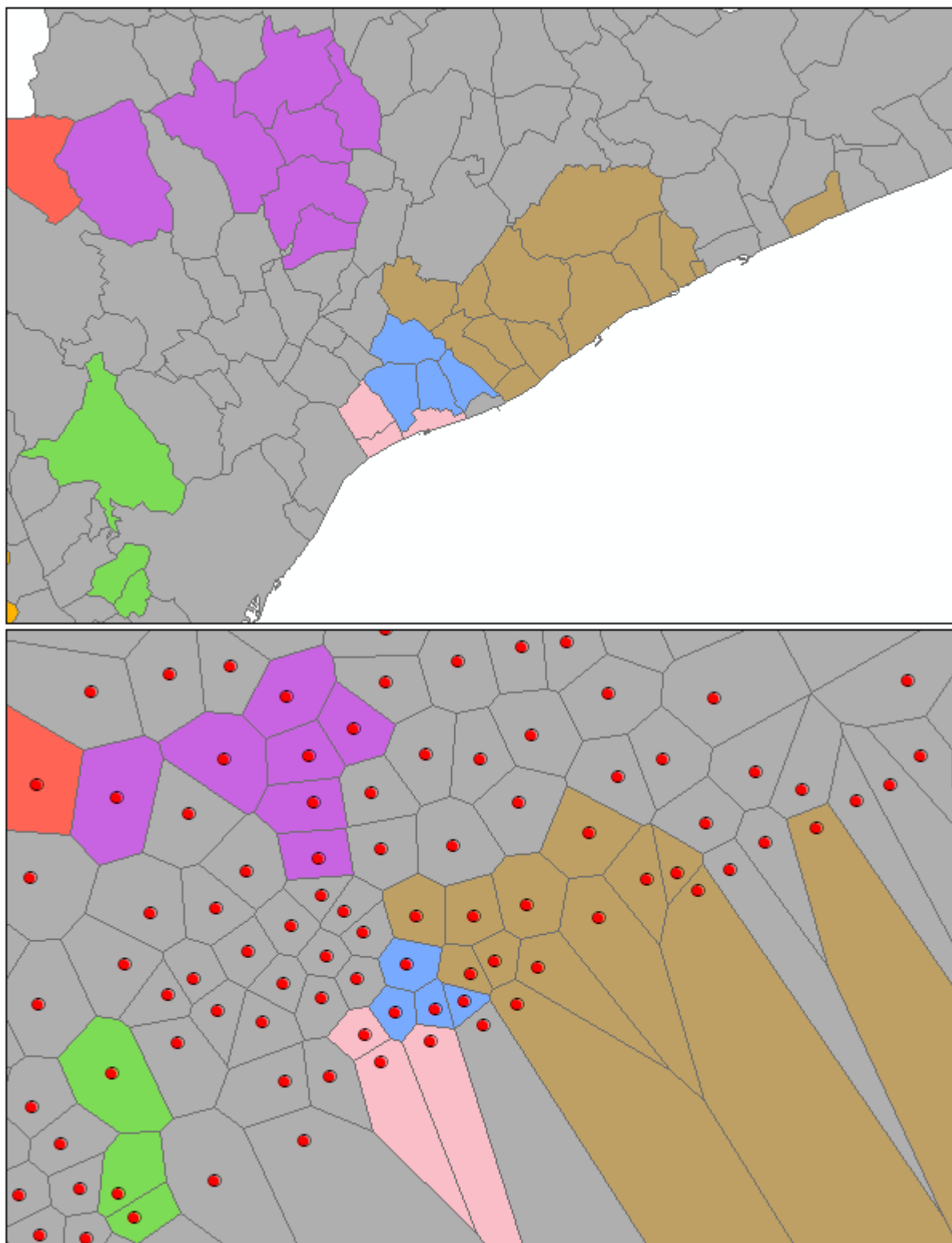


Рис. 2.7. Кластеризація полігональних та точкових об'єктів

Просторові обмеження для групування використовують алгоритм К-середніх. Мета цього алгоритму - розбити об'єкти так, щоб відмінності між об'єктами в групі (для всіх груп) були мінімальними. Так як алгоритм є NP-складним, для групування об'єктів використовується жадібна евристика. Жадібний

алгоритм завжди зводиться до локального мінімуму, але не завжди знаходить глобальний (оптимальний) мінімум.

Алгоритм К-середніх спочатку визначає початкові об'єкти, які використовуються для формування кожної групи. Відповідно число початкових об'єктів завжди дорівнює параметру Кількість груп. Перший початковий об'єкт вибирається довільно. При виборі залишилися початкових значень (хоча випадковий компонент також використовується) застосовується зважування, яке віддає перевагу об'єктам, найбільш віддаленим від існуючого набору початкових об'єктів (ця частина алгоритму називається К-середніх).

Після визначення початкових значень всі об'єкти призначаються найближчому початковому об'єкту (в просторі даних). Для кожного кластера об'єктів обчислюється центр даних, а кожен об'єкт призначається до найближчого центру. Процес обчислення центру даних для кожної групи і призначення об'єктів найближчого центру триває до стабілізації груп.

Мінімальна крона дерева. При вказівці просторового обмеження, щоб включати в групу суміжні або близькі об'єкти, інструмент спочатку формує граф зв'язності, що представляє сусідські відносини об'єктів. На основі графа зв'язності формується мінімальний кістяк дерева, який відображає просторові відносини об'єктів і схожість даних об'єктів. Об'єкти стають вузлами в мінімальній кроні дерева, пов'язаній зваженими ребрами. Вага кожного ребра пропорційна подібності об'єктів, які він з'єднує. Після створення мінімальної крони дерева гілка (ребро) дерева обрізається, після чого ми отримуємо два мінімальних графа дерева. Ребро для обрізання вибирається так, щоб мінімізувати розбіжність в отриманих групах, уникаючи при цьому отримання груп тільки з одним об'єктом. При кожній ітерації одна з мінімальної крони дерева ділиться цим процесом обрізки до отримання зазначеної кількості груп. Опублікований метод називається SKATER (просторовий кластерний аналіз з видаленням ребра дерева). Хоча на кожній ітерації вибирається гілка, що оптимізує схожість об'єктів в групі, не можна гарантувати, що кінцевий результат буде оптимальним.

2.4. Моделювання просторових відносин

Концептуалізація просторових відносин. Важлива відмінність між просторовою і традиційною (просторовою або непросторовою) статистикою полягає в тому, що просторова статистика інтегрує простір і просторові відносини безпосередньо в обчисленнях. Загальна концептуальна модель (концептуалізація) включає зворотню відстань, час у дорозі, фіксовану відстань, K найближчих сусідів і суміжність. Концептуальна модель (концептуалізація) просторових відносин залежить від того, який об'єкт аналізується. Якщо ви вимірюєте кластеризацію окремих видів рослин, що розмножуються насінням, ймовірно, зворотня відстань підходить найкраще. Однак, якщо ви оцінюєте географічний розподіл жителів передмість, що приїжджають на роботу в місто, час і вартість шляху будуть найкраще описувати просторові відносини. Для деяких видів аналізу, простір і час можуть мати менше значення, ніж абстрактні концепції, такі як знайомство з чим-небудь (чим більше знайоме щось, тим більшою функціональністю воно володіє) або просторова взаємодія (наприклад, між Нью-Йорком і Лос-Анджелесом відбувається значно більше телефонних дзвінків, ніж між Нью-Йорком і невеликим містом поруч з ним, таким як Покіпсі - можна припустити, що Лос-Анджелес і Нью-Йорк функціонально ближче).

Зворотня відстань, зворотня відстань в квадраті (опір)

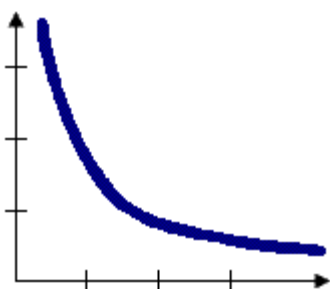


Рис. 2.8. Концептуалізація просторових відносин методом зворотної відстані

При використанні опцій зворотної відстані, концептуальна модель (концептуалізація) просторових відносин залежить від опору, або загасання в

залежності від відстані. Всі просторові об'єкти впливають на інші просторові об'єкти, але, зі збільшенням відстані, цей вплив знижується. У загальному випадку, якщо ви використовуєте концептуалізацію зворотніх відстаней, вам буде потрібно ввести значення діапазон відстаней або гранична відстань для скорочення кількості необхідних обчислень, особливо для великих наборів даних. Якщо ці значення не вказані, обчислюється порогове значення за замовчуванням. Можете ввести всі об'єкти в найближчу околицю, задавши нульове значення для параметра діапазону відстаней.

Зворотня евклідова відстань підходить для моделювання неперервних даних, наприклад змін температури. Зворотня Манхеттенська відстань найкраще працює, якщо в аналізі беруть участь розташування магазинів або інших міських підприємств, в тому випадку, якщо дані дорожньої мережі недоступні. Концептуальна модель при використанні опції Зворотні відстані в квадраті не відрізняється від моделі, що використовує зворотні відстані, за винятком випадків, коли ухил гостріше, і вплив сусідів спадає швидше і тільки найближчий сусід робить найбільш значущий вплив на обчислення для даного об'єкта (Рис. 2.9).

Діапазон відстаней (сфера впливу)

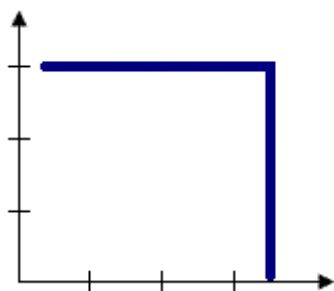


Рис. 2.9. Концептуалізація просторових відносин методом діапазону відстаней

Для таких інструментів як Аналіз гарячих точок (Hot Spot Analysis) діапазон фіксованих відстаней вибирається як концептуалізація за замовчуванням. За допомогою опції Діапазон фіксованих відстаней (Fixed Distance Band) вводиться модель «сфери впливу» або «ковзаючого вікна». Кожен просторовий об'єкт

аналізується в контексті навколишніх його об'єктів, розташованих на відстані, заданому параметром Діапазон відстаней або порогове відстань (Distance Band or Threshold Distance). Сусіди в межах заданої відстані мають однакову вагу. Просторові об'єкти, що знаходяться поза вказаної відстані, не впливають на обчислення (їх вага приймається рівною 0). Метод Діапазон фіксованих відстаней використовується, якщо треба оцінити статистичні параметри даних при певному (фіксованому) просторовому масштабі. Якщо ми вивчаємо закономірності щоденних поїздок в місто з передмістя і знаємо, що середня відстань поїздки становить 15 км, ви можете використовувати для аналізу фіксовану відстань в 15 км. (Рис. 2.10).

Зона індиферентності

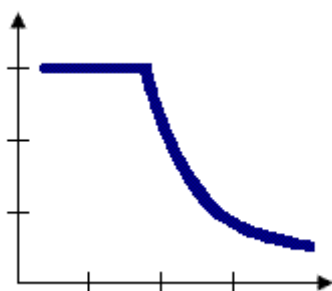


Рис. 2.10. Концептуалізація просторових відносин методом зон індиферентності

Зона індиферентності комбінує моделі Зворотня відстань і Діапазон фіксованих відстаней. Просторові об'єкти, розташовані в межах діапазону відстаней або порогової відстані, включаються в аналіз цільового просторового об'єкта. Після досягнення критичної відстані, рівень впливу (вага) швидко зменшується. Цей метод використовується, якщо потрібно мати фіксований масштаб аналізу, але не треба обмежуватися жорсткими рамками навколишніх об'єктів включених в обчислення для цільового об'єкта.

Концептуалізація відносин через суміжність полігонів першого порядку

Для класів полігональних об'єктів використовуються варіанти опцій СУМІСНІ РЕБРА або СУМІСНІ ВЕРШИНИ. В рамках методу СУМІСНІ РЕБРА

полігони, які мають загальне ребро (тобто спільний кордон), включаються до обчислення для цільового полігону. Полігони, які не мають загальні ребра, виключаються з обчислень. В рамках методу СУМІСНІ ВЕРШИНИ полігони, які мають загальне ребро і (або) вершину, включаються в обчислення для цільового полігону. Якщо частина двох полігонів перетинається, вони вважаються сусідами і включаються у відповідні обчислення. Ці концепції використовуються для опрацювання даних, що моделюють безперервні явища.

Найближча окіл K

Відносини в околиці можна також сформулювати наступним чином, що кожен просторовий об'єкт буде доступний в межах просторового контексту, що складається з певної кількості найближчих околів. Якщо K (число околів) дорівнює 8, в обчислення для цільового просторового об'єкта будуть включені вісім найближчих околиць. У тих місцях, де щільність просторових об'єктів висока, просторовий контекст аналізу буде менше. Відповідно, якщо щільність об'єктів невелика, просторовий контекст аналізу буде більше. Перевага цієї моделі просторових відносин полягає в тому, що для кожного цільового об'єкта буде задано кілька околиць, навіть якщо щільність просторових об'єктів у досліджуваній області варіює в широких межах. Цей метод можна застосувати під час побудови матриці просторових ваг (Матриця генерації просторових ваг).

Триангуляція Делоне (природне оточення)

Триангуляція Делоне створює окіл методом побудови трикутників Вороного на основі точкових об'єктів або за центроїдами просторових об'єктів так, щоб кожна точка / центроїд ставала вузлом (вершиною) трикутника. Вузли, з'єднані ребрами трикутників, розглядаються як околиці. Використання триангуляції Делоне дозволяє забезпечити для кожного просторового об'єкта наявність хоча б одного околу, навіть якщо в даних присутні острова і/або щільність просторових об'єктів змінюється в широкому діапазоні (Рис. 2.11).

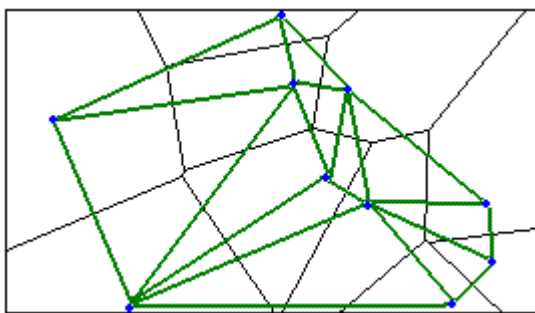


Рис. 2.11. Триангуляція Делоне

Просторово-часове вікно

За допомогою цієї концептуалізації можна визначити відношення об'єктів як в просторовому (фіксована відстань), так і в часовому (фіксований інтервал) вікні. Ця опція доступна при створенні файлу матриці просторових ваг. Опціями для визначення інтервалу є входні дані із вказанням дати або часу події або періоду між відповідними подіями. З такою концептуалізацією об'єкти будуть сусідами, якщо вони розташовані на певній відстані і в межах заданого тимчасового інтервалу цільового об'єкта. Цей тип концептуалізації може використовуватися для інструменту Аналіз гарячих точок, щоб виявити гарячі точки в просторі-часі.

Рекомендації щодо вибору концептуалізації просторових відносин:

Параметр визначення просторових взаємовідносин повинен відображати внутрішні відносини між просторовими об'єктами, які аналізуються. На вибір можуть впливати характеристики наявних даних.

Методи обернено-зваженої відстані, наприклад найкраще підходять для безперервних даних або для моделювання процесів, в яких два об'єкти зближені в просторі настільки, що з великою ймовірністю впливають один на одного. При такій просторовій концептуалізації, кожен об'єкт є потенційною околицею будь-якого іншого об'єкта, і при використанні великих наборів даних кількість обчислень може стати значною. Завжди слід намагатися включити значення Діапазон відстаней або порогове значення при використанні концептуалізації обернено-зваженої відстані. Це особливо важливо для великих наборів даних. Якщо не задавати параметр Діапазон відстаней або порогове значення, гранична відстань буде обчислена автоматично, але вона може виявитися не цілком відповідною відстані, яка повинна враховуватися в аналізі. За замовчуванням

гранична відстань дорівнює мінімальній відстані, яка гарантує, що для кожного об'єкта є хоча б одна околиця.

Метод фіксованого діапазону відстаней добре працює з полігональними даними, що мають різні розміри полігонів (великі полігони по краях досліджуваної області і дуже маленькі полігони в центрі досліджуваної області, наприклад), якщо вам потрібно забезпечити узгоджений масштаб аналізу.

Концептуалізація зони індиферентності добре працює при правильно підібраній фіксованій відстані, але занадто різкі кордони відносин околиць заважають точному поданню даних. Концептуалізація зони індиферентності розглядає кожен просторовий об'єкт як околицю іншого об'єкта. Отже, ця опція не підходить для великих наборів даних, оскільки зазначені значення діапазону відстаней або порогового відстані не обмежують число околиць, а тільки визначають, де інтенсивність просторових відносин починає знижуватися.

Концептуалізація суміжності полігонів ефективна, коли полігони мають однакові розміри і однотипне розподіл, а просторові відносини є функцією близькості полігонів (якщо два полігони мають спільний кордон, просторове взаємодія між ними зростає).

Опція К найближчих сусідів ефективна, якщо необхідно задати мінімальну кількість околиць для аналізу. Якщо значення, пов'язані з просторовими об'єктами, асиметричні (не мають нормального розподілу), важливо, щоб кожен об'єкт оцінювався в контексті як мінімум восьми (або близько того) околиць. Якщо розподіл даних в досліджуваному районі неоднорідний, тобто, деякі просторові об'єкти розташовані дуже далеко від всіх інших об'єктів, цей метод працює дуже добре. Однак треба зауважити, що просторовий контекст аналізу змінюється в залежності від зміни щільності використовуваних просторових об'єктів. Якщо фіксований масштаб аналізу має менше значення, ніж фіксоване число околиць, можна використовувати метод К найближчих сусідів.

Деякі аналітики розглядають триангуляцію Делоне в якості способу створення природних околів для набору просторових об'єктів. Цей метод є ефективним, якщо дані містять острівні полігони (ізолювані полігони, які не мають

спільних меж з іншими полігонами), або в тих випадках, коли просторові об'єкти розподілені надто нерівномірно. Цю опцію не слід використовувати для наборів даних з однаковими об'єктами. Так само, як і метод К найближчих сусідів, триангуляція Делоне забезпечує для кожного просторового об'єкта наявність мінімум однієї околиці, при цьому на основі розподілу даних визначається кількість околиць, яку може бути присвоєно кожному об'єкту.

Опції Просторово-часове вікно (Space-Time Window) дозволяє визначити відносини об'єктів як з точки зору просторової, так і часової близькості. Ця опція використовується для визначення гарячих точок в просторі-часі або формування груп, членство в яких визначається близькістю в просторі і часі.

2.5. Метод просторово-часового кубу для аналізу геопросторових даних

Останнім часом використання великих масивів даних, що розосереджені не лише по значній території, але і у великому часовому інтервалі, потребує відповідних технологій для прийняття правильного рішення. Одним з варіантів вирішення даного питання є дослідження, аналіз та прогнозування події за допомогою побудови просторово-часового кубу. Вперше використання просторово-часового кубу запропонував Т. Хейгерstrand на початку 70-х років [271], можливості якого він описав у своїй праці «What about people in regional science?». Але до активного розвитку геоінформаційних систем (ГІС) його використання було обмежене. Лише в 2000 – х роках з'являються роботи з використання просторово-часового кубу в ГІС. У цих роботах представлялися нові можливості використання просторово-часового кубу за допомогою ГІС, в тому числі включаючи дослідження землетрусів [197, 252, 294].

Наступними кроками у використанні методу просторово-часового кубу було його застосування в інтелектуальному аналізі даних різноманітної природи: аналіз злочинів, вивчення інфраструктури, аналіз поведінки тварин, візуалізації руху людини, дослідження залежностей зміни погодних умов в часі [222, 258, 407]. В Україні лише деякі роботи лише частково розв'язують проблему використання просторово-часового кубу, але не розкривають її повністю [186]. Таким чином,

моделювання великих масивів багатовимірних даних, їх аналіз та використання для потреб прийняття рішень у різних сферах є важливим завданням, що потребує вирішення.

Просторово-часовий куб – це техніка 3D-візуалізації, призначена для одночасного представлення просторових та часових характеристик руху. Згідно з цим, точки траєкторій відображаються у тривимірному просторі, де вертикальна вісь зазвичай виражає час [186].

На початку 70-х років Т. Хейгерстранд [271] розробив графічний погляд на час як додатковий просторовий вимір. Він запропонував тривимірну діаграму, так званий просторово-часовий куб, щоб показати життєві історії людей і те, як люди взаємодіють у просторі та часі. Площина куба являє собою двовимірний географічний простір, а висота куба - часовий вимір. У той час, коли була представлена концепція, варіанти створення такої графіки обмежувались ручними методами, що слугувало очевидним обмеженням підходу, оскільки створення кожної нової діаграми виявилось трудомістким заняттям. Сьогодні ж, коли сучасні комп'ютерні технології мають набагато кращі можливості для візуалізації даних, графічний погляд Т. Хейгерстранда отримує друге життя та активно застосовується у ГІС.

Використання просторово-часового кубу потребує просторові та часові дані для аналізу певних подій. Прикладами таких подій є землетруси, дорожньо-транспортні пригоди, випадки захворювань або спостереження рідкісних тварин [4].

Т. Хейгерстранд пропонував застосовувати просторово-часовий куб до даних про рух об'єктів, тобто про зміни просторових місць з прив'язкою до часу. У даній роботі автори пропонують застосовувати концепцію Т. Хейгерстранда до іншого типу даних, а саме до аналізу мережевих подій. Окрім цього використання просторово-часового кубу дає змогу відповісти на 3 питання Пьюке, поставлені до просторово-часових даних [236]:

- коли + де → що: опис об'єктів або набір об'єктів, які присутні в певному місці або наборі місцеположень за певний час або проміжок часу;

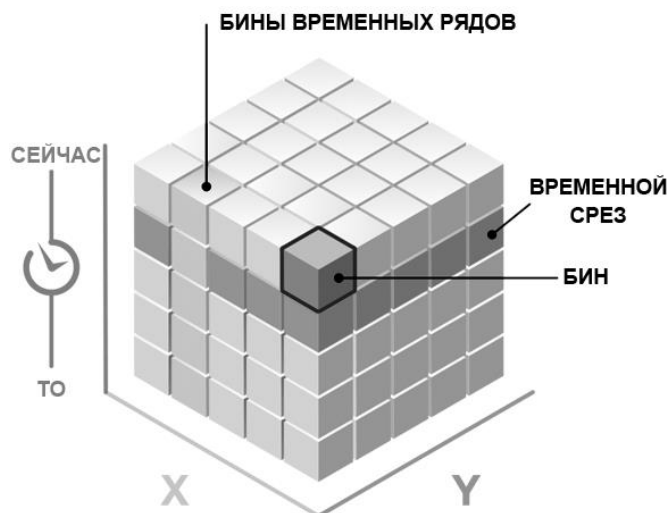
- коли + що → де: опис місця розташування або набір місць розташування, зайнятих певним об'єктом або набором об'єктів у певний час або проміжку часу;
- де + що → коли: опис конкретного часу або проміжку часу, коли певний об'єкт або набір об'єктів займали певне місце або набір місць розташування.

Для проведення дослідження використовуються дані надані компанією Vodafone, котрі мають просторову та часову прив'язку, а також певну атрибутивну інформацію. Опрацьована база даних нараховує 150 тисяч дзвінків, повідомлень та виходів у інтернет, з найрізноманітніших пристроїв та від різних абонентів. Усі «події» зосереджені практично у західних областях України.

За для досягнення поставлених цілей, автори використовують набір інструментів Поглибленого аналізу просторово-часових закономірностей у програмному забезпеченні ArcMap 10.5. Даний набір інструментів містить статистичні інструменти для аналізу розподілу даних і виявлення закономірностей в контексті простору-часу. У набір входять інструменти: Створити куб Простір-Час та Аналіз виникнення гарячих точок.

Інструмент Створити куб Простір-Час на основі набору точкових даних будує тривимірну кубічну структуру (в форматі netCDF) для подальшого аналізу. Інструмент Аналіз виникнення гарячих точок використовує цей куб в якості вхідних даних і на основі нього виявляє статистично значущі тренди гарячих і холодних точок у часі. Інструмент Аналіз виникнення гарячих точок можна використовувати при обробці даних про правопорушення або спалахи епідемій для виявлення нових, зростаючих, спорадичних або постійних гарячих точок, досліджуючи серії тимчасових інтервалів [243].

Інструмент Створити куб Простір-Час використовує точкові об'єкти з часовим атрибутом і структурує їх в куб формату netCDF, агрегуючи точки в біни простору-часу (Рис. 2.12). Для значень бінів виконується підрахунок кількості точок і визначається наявність трендів в часі в кожному місцезнаходження з



використанням статистики Манна-Кенделла. [245]

Рис. 2.12. Відображення бінів у просторово-часовому кубі [243].

Структура куба містить рядки, стовпці і часові кроки. Перемноживши кількість рядків на число стовпців і на число часових кроків, отримуємо сумарне число бінів у кубі. Рядки і стовпці визначають просторовий екстент куба, а часові кроки визначають часовий екстент. Місця розташування з даними (біни в просторовому вимірі) - це області, розташовані в межах екстента, де протягом усього часового інтервалу зафіксовано хоча б одна точкова подія. [245]

Для більшості інструментів аналізу тільки місце розташування з даними хоча б для одного часового кроку буде включено в аналіз, при цьому аналіз буде проводитися для всіх часових кроків. Нульову кількість буде присвоєно біну, що не містить жодної точки, якщо у відповідному місцезнаходженні хоча б один часовий крок буде містити хоча б одну точку. Інформація про відсоток бінів з нульовою кількістю, пов'язаних з місцями розташування, для яких є дані хоча б для одного часового кроку, буде відображена в повідомленні, як величина розрідженості. [245]

Вхідні об'єкти повинні бути виключно точковими об'єктами, які представляють дані подій, наприклад, точки злочинів або пожеж, точок спалаху захворювання або дорожніх пригод, у нашому випадку подія, що пов'язана із вхідними чи вихідними дзвінками та повідомленнями, а також виходами користувачів у інтернет. У кожної точки повинна бути прив'язка до дати. Поле зі значенням часу для події повинно мати тип «Date». Ще однією умовою даного інструменту є наявність, як мінімум, 60 точок з різними значеннями часу. Даний інструмент припинить роботу, якщо в ході роботи буде створено куб з числом елементів, що перевищує два мільярди елементів (бінів). Для точного вимірювання відстаней цьому інструменту потрібні дані у прямокутній системі координат.

Разом із створеним просторово-часовим кубом інструмент надає вихідне повідомлення, у кінці якого знаходиться інформація про загальний тренд в даних. Тренд обчислюється на основі аналізу часових рядів в просторі. Основне питання, на яке відповідає аналіз - кількість подій, які представлені вхідними точками, збільшується або зменшується з плином часу? Щоб отримати відповідь на це запитання, значення кількості точок за місцем розташування групуються в тимчасові ряди і аналізуються з використанням статистики Манна-Кендалла [244].

Непараметричний тест Манна-Кендалла зазвичай використовується для виявлення тенденцій в серії даних. Нульова гіпотеза H_0 полягає в тому, що вибірка даних хронологічно впорядкована, незалежна і ідентично розподілена. Альтернативна гіпотеза H_A полягає в тому, що дані відповідають однозначному розподілу. Статистика тесту Манна-Кендалла розраховується за:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \operatorname{sgn}(X_j - X_k), \quad (2.27)$$

де

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}. \quad (2.28)$$

Середнє $S \in E[S] = 0$ та варіація $\sigma^2 \in$

$$\sigma^2 = \{n(n-1)(2n+5) - \sum_{j=1}^p t_j(t_j-1)(2t_j+5)\}/18, \quad (2.29)$$

де p - число зв'язаних груп в наборі даних, t_j - кількість точок даних в j зв'язаній групі. Статистика S приблизно нормально розподілена за умови, що використовується наступне Z -перетворення:

$$Z = \begin{cases} \frac{S-1}{\sigma} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sigma} & \text{if } S < 0 \end{cases}, \quad (2.30)$$

Статистика S є дуже близькою до τ Кендела як показано нижче:

$$\tau = \frac{S}{D}, \quad (2.31)$$

де

$$D = \left[\frac{1}{2} n(n-1) - \frac{1}{2} \sum_{j=1}^p t_j(t_j-1) \right]^{1/2} \left[\frac{1}{2} n(n-1) \right]^{1/2}. \quad (2.32)$$

Горизонтальні розміри біна за замовчуванням розраховуються наступним чином [13]:

- Обчислюється довжина більш довгої сторони екстента вхідних об'єктів (максимальний екстент).
- Далі розмір біна розраховується або як значення максимального екстента, розділене на 100, або за алгоритмом, заснованому на просторовому розподілі вхідних об'єктів (береться більше значення).

Інтервал часового кроку за замовчуванням розраховується з використанням двох різних алгоритмів, що визначають оптимальну кількість і ширину інтервалів часового кроку. Мінімальна кількість більше 10, отримане в результаті розрахунків з використанням двох алгоритмів, використовується як значення за замовчуванням для інтервалу часового кроку. Якщо в результаті обох алгоритмів отримано число менше 10, значення за замовчуванням для інтервалу часового кроку встановлюється на 10.

Отриманий набір даних Vodafone знаходиться у часовому відрізку з 1 червня 2017 року по 31 серпня 2017 року. За для зручності аналізу було використано 5-ти

денний часовий крок. У наслідок чого інструментом було побудовано куб висотою у 19 бінів (Рис. 2.13).

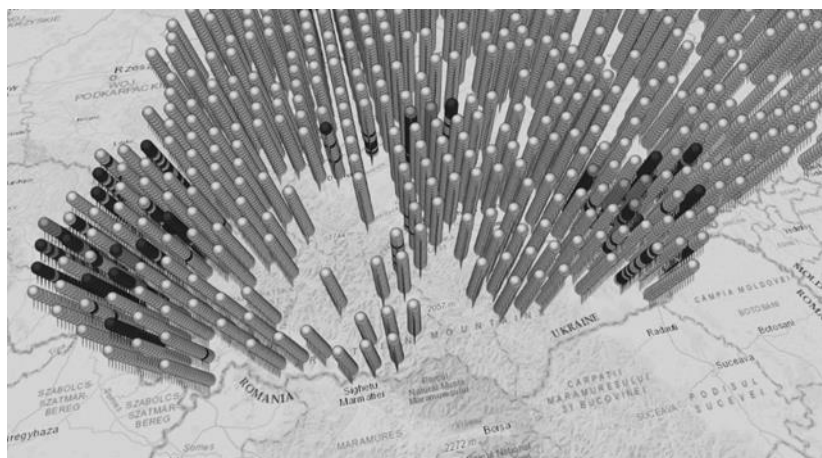


Рис. 2.13. 3D візуалізація просторово-часового кубу, західних регіонів України

Інструмент Аналіз виникнення гарячих точок ідентифікує тренди в кластері щільності точок (підрахунків) або поля сум в просторово-часовому кубі, створеному за допомогою інструменту Створити куб простір-час. До категорій холодних і гарячих точок відносяться наступні характеристики [14]: нова, послідовна, зростаюча, постійна, спадна, спорадична, що коливається і історична.

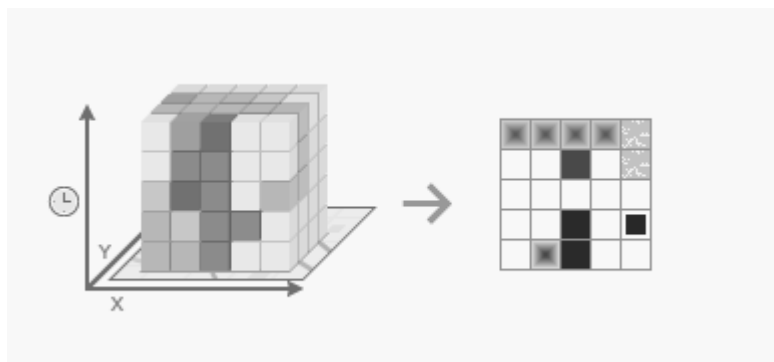


Рис. 2.14. Трансформація просторово-часового кубу для аналізу через гарячі точки [245].

Для цього інструменту необхідний файл netCDF, створений інструментом Створити куб Простір-час.

Кожен бін у просторово-часовому кубі містить значення LOCATION_ID, time_step_ID, COUNT і, можливо, поля підсумовування. Набір бінів, пов'язаний з

одним і тим же місцем розташування, має однаковий ідентифікатор місця розташування і являє собою тимчасовий ряд. Набір бінів, пов'язаний з одним і тим же тимчасовим інтервалом, має однаковий ідентифікатор часового кроку і являє собою часовий зріз. Значення кількості в кожному біні представляє число точок, які присутні на певній території та певному часовому інтервалі.

Даний інструмент аналізує варіабельність у Вхідному кубі простору-часу netCDF, використовуючи просторово-часову інтерпретацію статистики Getis-Ord G_i^{**} , при цьому обчислюється значення кожного біна, яке зіставляється зі значеннями в сусідніх бінах.

Вихідні об'єкти додаються до Таблиці змісту і являють узагальнений результат просторово-часового аналізу для всіх проаналізованих місцеположень. Крім створення класу Вихідних об'єктів, зведені результати аналізу записуються в вікно Результати. Основним результатом використання інструменту Аналіз виникнення гарячих точок є створення карти, за допомогою якої надалі можливе дешифрування та знаходження гарячих точок (Рис. 4).

Серед інструментів поглибленого аналізу просторово-часових закономірностей є інструмент: Аналіз локальних викидів, який так само як і попередній дозволяє дешифрувати важливі показники у Просторово-часовому кубі (Рис.5). Для ідентифікації кластерів та викидів у просторово-часовому кубі, даний інструмент використовує просторово-часову інтерпретацію статистики Anselin Локальний індекс Морана I, при цьому значення кожного біна зіставляється зі значеннями в сусідніх бінах (рис. 2.15, 2.16).

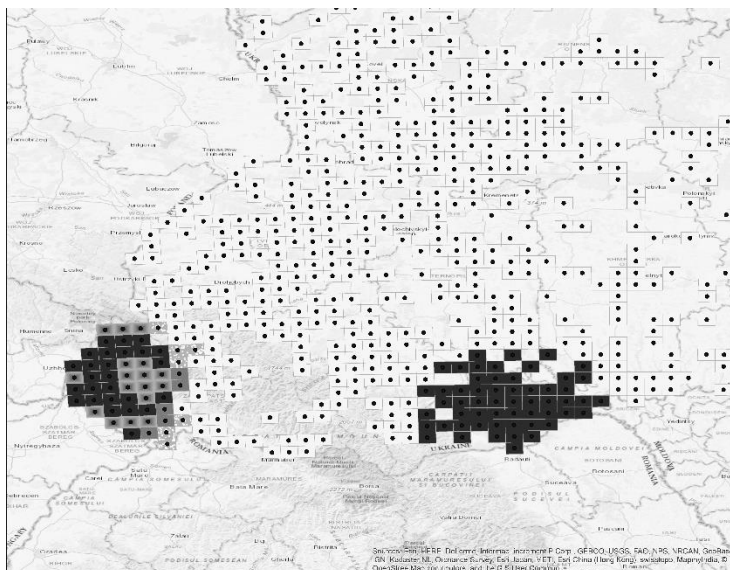


Рис. 2.15. Карта виникнення гарячих точок, створена в результаті аналізу просторово-часового кубу

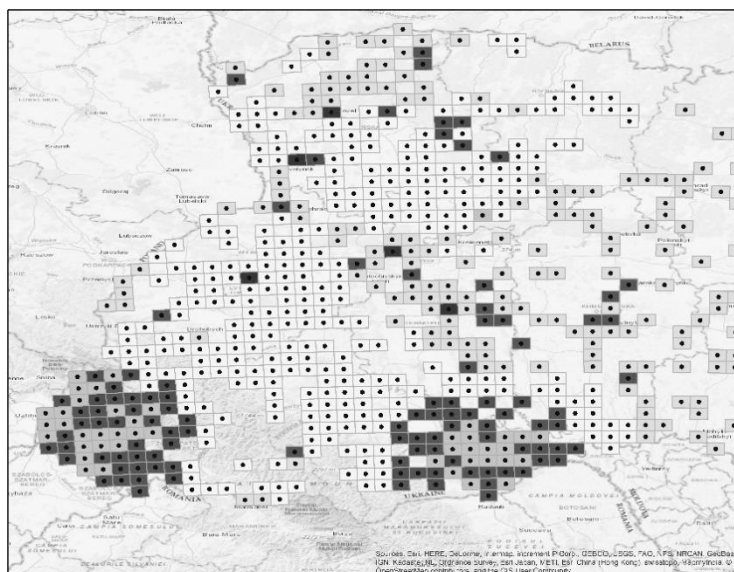


Рис. 2.16. Розподіл локальних викидів

Щоб ідентифікувати біни, які будуть включені в кожну околицю аналізу, інструмент спочатку виявляє біни, які потрапляють в задані параметри аналізу. Потім для кожного біна визначаються біни, розташовані в тому самому розташуванні, але в межах N попередніх тимчасових кроків, де N - Тимчасовий крок околиці, зазначений у вхідних параметрах.

Біни з високими значеннями локальних викидів містять аномальні зміни у поведінці користувачів, які можуть мати різну природу як позитивного, так і

негативного характеру. Разом з використанням класифікаторів та соціальних каналів розповсюдження новин, вони можуть бути ідентифіковані та передані до відповідних державних органів та служб.

Висновки до розділу

В другому розділі визначено, що процедури ІАГД реалізуються на основі використання апарату геоінформаційних систем, який передбачає спеціальні процедури організації зберігання та обробки геоданих. Зберігання геопросторових даних для розв'язання проблеми оцінювання індикаторів сталого розвитку організовано на основі векторних та растрових моделей організації даних в базах геоданих. Встановлено особливості застосування ІАГД під час обробки даних на основі топологічних та нетопологічних векторних та растрових моделей.

Встановлено, що базові методи ІАГД визначаються типом розподілу даних. На основі гіпотезою імовірності оцінки виникнення аномальних значень на обмеженій просторовій відстані. Для визначення цих параметрів запропоновано використання методів просторової автокореляції між геопросторовими об'єктами (Глобальний індекс І Морана), показник середнього найближчого сусідства, локальний індекс Морана. Для цілей просторової кластеризації запропоновано проведення процедури аналізу гарячих точок (Getis-Ord G_i^*), в якому ефективність угруповання об'єктів вимірюється за допомогою псевдо-F-статистики Калінскі-Харабаза. Просторова кластеризація в цьому випадку визначається просторовими обмеженнями топології. Запропоновано підходи до моделювання просторових відносин методами зворотньої відстані, діапазонів відстаней, зон індиферентності, суміжності та сусідства.

Сукупність цих методів вперше застосовано на прикладі аналізу великих геопросторових даних з використанням побудови моделі просторово-часового кубу – техніки 3D-візуалізації та аналізу даних, призначеної для одночасного представлення просторових та часових характеристик. Згідно з цим, точки

траекторій відображаються у тривимірному просторі, де вертикальна вісь зазвичай відповідає за час.

Приклад використання просторово-часового кубу для аналізу даних подій у телекомунікаційних мережах Vodafone на території західної частини України надає змогу більш ефективно використовувати отримані дані, перш за все у безпекових цілях, що буде корисно урядовим організаціям для швидкого виявлення або попередження небезпечних ситуацій (таких як тероризм, надзвичайні явища, катастрофи чи інше).

Для цілей аналізу було використано набір даних обсягом 1,5 млн. записів про події у мережі. В результаті було запропоновано використанням методу гарячих точок та статистики Манна-Кендела для визначення закономірності та аномальні статистичні викиди у розподілі даних, які характеризувалися просторовою та часовою динамікою. Було побудовано візуалізацію даних у вигляді тривимірних карт та встановлено кількісні зв'язки між суспільно значущими подіями та активністю абонентів.

Надалі за допомогою просторово-часового кубу на основі даних мобільних операторів можна аналізувати статистичні викиди в активності абонентів у дзвінках або підключенню людей до інтернет з прив'язкою до певної території, що дасть змогу ідентифікувати певні аномалії та відповідно реагувати на ситуацію.

РОЗДІЛ 3. МЕТОДИ КЛАСТЕРИЗАЦІЇ ГЕОПРОСТОРОВИХ ДАНИХ

3.1. Огляд класичних методів класифікації багатовимірних даних

3.1.1 Постановка завдання класифікації багатовимірних даних

В даний час в прикладній статистиці склалася система методів автоматичної класифікації [1, 2, 3]. У загальному випадку вона може бути охарактеризована наступним чином.

Позначимо всю кількість об'єктів, що класифікуються символом:

$O = \{o_1, \dots, o_N\}$, де

O_i - i -й об'єкт,

N – кількість об'єктів.

Метою класифікації є отримання деякої наперед заданої або не заданої кількості груп об'єктів (класів об'єктів). У межах кожного класу об'єкти повинні бути максимально «схожі» один на одного в деякому сенсі, однорідні, а об'єкти з різних класів - максимально «відрізнитися».

Будемо називати u -м класом підмножин об'єктів

$$S_i = \{o_{ij} \in O \mid j \in 1, \dots, N_i\} O, \quad \text{де}$$

$N_i = |S_i| \leq N$ - кількість об'єктів i -го класу (потужність i -го класу).

Під системою класів, отримання якої є метою класифікації, будемо розуміти множину S з наступними властивостями:

1. $S = \{S_i \mid i \in 1, \dots, K\}$, тобто S складається із K класів.

2. $S_i \cap S_j = \emptyset \quad \forall i, j \in \{1, \dots, K\}$, де \emptyset - порожня множина.

3. $S_i \neq \emptyset$, т. е. $|S_i| > 0 \quad \forall i \in \{1, \dots, K\}$

$$4. \bigcup_{i=1}^K S_i = O, \quad \sum_{i=1}^K N_i = N \quad . \quad (3.1)$$

Задати систему класів можна IV -мірним вектором приналежності об'єктів до одного з класів:

$$v = (v_1, \dots, v_N),$$

$$v_i \in \{1, \dots, K\} : v_i = j \Leftrightarrow v_i \in S_j, i \in \{1, \dots, N\}, j \in \{1, \dots, K\}$$

Результатом знаходження об'єктів в просторі ознак є одним з двох представлень:

- представлень вихідних об'єктів у вигляді матриці «об'єкт-ознака» («об'єкт-властивість»), що відображає вимір M ознак на N об'єктах і містить N рядків і M стовпців:

$$X = \begin{pmatrix} o_1 \\ \dots \\ o_N \end{pmatrix} = (o^{(1)}, \dots, o^{(M)}) = \begin{pmatrix} o_1^{(1)} & \dots & o_1^{(j)} & \dots & o_1^{(M)} \\ o_2^{(1)} & \dots & o_2^{(j)} & \dots & o_2^{(M)} \\ \dots & \dots & \dots & \dots & \dots \\ o_{N-1}^{(1)} & \dots & o_{N-1}^{(j)} & \dots & o_{N-1}^{(M)} \\ o_N^{(1)} & \dots & o_N^{(j)} & \dots & o_N^{(M)} \end{pmatrix}, \quad (3.2)$$

де $o_i = (o_i^{(1)}, \dots, o_i^{(M)})$ - i -й об'єкт в M - вимірному просторі ознак, $o^{(j)}$ - j - й ознак, $o^{(j)} = (o_1^{(j)}, \dots, o_N^{(j)})^T$, $o_i^{(j)}$ - значення j -ї ознаки на i -му об'єкті, $i \in \{1, \dots, N\}$, $j \in \{1, \dots, M\}$; представлення вихідних об'єктів у вигляді матриці «об'єкт-об'єкт», що відображає результат зіставлення об'єктів в просторі ознак між собою і містить N рядків і стовпців:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1N} \\ a_{21} & \dots & a_{2j} & \dots & a_{2N} \\ \dots & \dots & \dots & \dots & \dots \\ a_{N-1} & \dots & a_{N-1} & \dots & a_{N-1N} \\ a_{N1} & \dots & a_{Nj} & \dots & a_{NN} \end{pmatrix}, \quad (3.3)$$

де a_{ij} - результат зіставлення i - го і j - го об'єктів, $i, j \in \{1, \dots, N\}$.

Таблиця 3.1

Систематизація завдань класифікації

Апріорні відомості про класи (генеральні сукупності)	Попередня вибіркова інформація	
	Немає інформації	Є навчальні вибірки
Деякі найзагальніші припущення про закон розподілу досліджуваного вектора: гладкість,	Класифікація без навчання: кластер-аналіз, таксономія, розпізнавання образів «без	Непараметричні методи дискримінантного аналізу

зосередженість в середині обмеженої області і т.п.	вчителя», ієрархічні класифікації	
Розрізняє генеральні сукупності задані у вигляді параметричного сімейства законів розподілу ймовірностей (параметри невідомі)	Інтерпретація досліджуваної генеральної сукупності як суміші декількох генеральних сукупностей. «Розщеплення» цієї суміші за допомогою методів оцінювання невідомих параметрів	Параметричні методи дискримінантного аналізу
Розрізняє генеральні сукупності задані однозначним описом відповідних законів	Класифікація при повністю описаних класах: розрізнення статистичних гіпотез	Навчальні вибірки не потрібні

При класифікації багатовимірних просторових даних наявність навчальних вибірок або повністю заданих законів розподілу є поодинокими випадками. Тому нижче розглядаються тільки методи розщеплення суміші розподілів і методи класифікації без навчання.

3.1.2 Модель суміші розподілів

Модель суміші розподілів для класифікації багатовимірних просторових даних можна використовувати в припущенні, що розподіли об'єктів всередині кожного класу відбуваються за нормальним законом. Вихідною інформацією для параметричної класифікації є подання даних таблиці «об'єкт-ознака». Модель суміші розподілів в найбільш простому випадку кінцевого числа класів M і однотипності компонентів записується у вигляді:

$$f_0(x) = \sum_{i=1}^M \rho_i f(x, \theta), \quad (3.4)$$

де $f_0(x)$ - щільність генеральної сукупності,

M - число компонент суміші,

p_i – вірогідність появи i -ї компоненти суміші,

$f(x_i, \theta_i)$ – щільність i -ї компоненти суміші,

θ_i – вектор параметрів для i -ї компоненти суміші (наприклад, для одновимірного нормального розподілу $\theta_i = (\mu_i, \sigma_i)$)

Модель суміші розподілів стосовно завдань класифікації має на увазі, що i -й клас повністю характеризується i -ю компонентою суміші $f(x_i, \theta_i)$ і вірогідністю її появи p_i . Завдання класифікації об'єктів полягає у визначенні, в рамках якого з класів поява даного об'єкта найбільш вірогідна. Іншими словами, класифікація здійснюється за правилом

$$o_i \in S_j \Leftrightarrow p_j f(o_j, \theta_j) = \max\{p_l f(o_l, \theta_l) | l \in 1, \dots, M\}. \quad (3.5)$$

Найбільш складним етапом при класифікації на основі моделі суміші розподілів є процедура ідентифікації суміші, тобто алгоритм отримання числа класів M і оцінок для p_i , та θ_i , які необхідні для побудови вирішального правила. Не всі суміші ідентифікуються, тобто не для всіх типів розподілу можна знайти єдині оцінки M , p_i , и θ_i . Наприклад, суміш нормальних розподілів ідентифікується, а суміш рівномірних – ні.

Існують різні підходи до оцінювання з множини об'єктів параметрів суміші, найбільш поширеним з яких є ЕМ-алгоритм. Назва ЕМ-алгоритм походить від англійського скорочення Estimation (оцінювання) і Maximization (максимізація). Цей метод для фіксованого числа класів (елементів суміші) K дозволяє визначати оцінки параметрів суміші $p_i, \theta_i, i \in \{1, \dots, K\}$ шляхом багаторазового знаходження чергових наближень до оцінок (крок Estimation) і максимізації з урахуванням наближень логарифмічної функції правдоподібності представленої в зручному для цього вигляді (крок Maximization).

$$L(p, \theta, K) = \ln\left(\prod_{i=1}^N f_0(o_i)\right) = \sum_{i=1}^N \ln f_0(o_i) = \sum_{i=1}^N \ln\left(\sum_{j=1}^K p_j f_j(o_i, \theta_j)\right). \quad (3.6)$$

ЕМ-алгоритм не дозволяє визначити оцінку кількості класів K . Кількість класів може бути знайдено стандартним способом шляхом завдання K_{\min} і K_{\max} , застосування ЕМ-алгоритму $\forall K \in [K_{\min}, K_{\max}]$, обчислення значення функції

правдоподібності $L(p, \theta, K)$ при підстановці знайдених оцінок і аналізу графіка $L(K)$.

Для визначення числа класів $K_{\text{ист}}$ існують і формальні процедури, одна з яких зводиться до перевірки серії гіпотез виду

$$\begin{aligned} H_0: K_{\text{ист}} &= K, \\ H_1: K_{\text{ист}} &= K = 1. \end{aligned}$$

Починаючи з $K = K_{\min}$

Нижче буде розглянуто різновид ЕМ-алгоритму (NEM-алгоритм), що враховує при знаходженні параметрів, крім атрибутивного, і географічний простір.

3.1.3 Методи класифікації без навчання

Методи класифікації без навчання можна поділити на два великі класи алгоритмів:

- методи, засновані на описі класів ядрами (паралельні і послідовні);
- ієрархічні процедури (агломеративні і дивізійні).

Зазначені класи процедур спираються на поняття відстані між векторами в багатовимірному просторі і відстані між класами об'єктів. Відстань між об'єктами в географічному просторі позначимо символом в просторі ознак - символом d :

$$g, d: O \times O \rightarrow R.$$

Відстань між об'єктами в просторі показників характеризує подібність або відмінність об'єктів між собою [3]. Кожному типу шкал, в якому виміряні ознаки, відповідає свій спосіб розрахунку відстані. Розглянемо лише деякі види завдання відстаней між об'єктами для кількісних, рангових або номінальних ознак.

Для розрахунку відстані на M кількісних ознаках існує найбільш загальне співвідношення, назване метрикою махаланобіського типу:

$$d_{muh}(o_i, o_j) = \sqrt{(o_i - o_j) \times \Lambda^{-1} \times (o_i - o_j)^T}, \quad (3.7)$$

Λ - ковариаційна матриця генеральної сукупності,

Λ - матриця вагових коефіцієнтів (сімеґрічна, невід'ємно визначена, найчастіше обирається діагональною, тобто

$$\Lambda = \text{diag}(\omega_1, \dots, \omega_M),$$

$o_i = (o_i^1, \dots, o_i^M)$, $o_i^{(x)}$, значення показника $x \in \{1, \dots, M\}$ на i -му об'єкті,

$$d_{ent}(o_i, o_j) = \sum_{k=1}^M I(o_i^{(k)} \neq o_j^{(k)}), \quad de \quad (3.8)$$

$$I(o_i^{(k)} \neq o_j^{(k)}) = \begin{cases} 1, & o_i^{(k)} \neq o_j^{(k)}, \\ 0, & o_i^{(k)} = o_j^{(k)}, \end{cases} \quad (3.9)$$

$$\text{Відповідно, } d_{ent}(o_i, o_j) \in \{0, \dots, M\}. \quad (3.10)$$

Для отримання повного переліку способів визначення відстаней між об'єктами в просторі ознак можна звернутися до спеціальної літератури по методам аналізу даних [232].

Нехай

$S = \{S_1, \dots, S_K\}$ — система класів,

$D : S \times S \rightarrow R$ - функція відстані між класами, визначена на всіх парах класів з S ,

$S_i = \{o_{i1}, \dots, o_{iN}\}$ - i -й клас, N - число об'єктів в i -му класі, $i \in \{1, \dots, K\}$

Нижче наведені найбільш відомі види відстані між класами (способи задання функції D).

1. Метод ближнього сусіда:

$$D_{min}(S_i, S_j) = \min\{d(o_{ix}, o_{jx}) | o_{ix} \in S_i, o_{jx} \in S_j\}. \quad (3.11)$$

2. Метод ближнього сусіда:

$$D_{min}^*(S_i, S_j) = (N_i + N_j) \times \min\{d(o_{ix}, o_{jx}) | o_{ix} \in S_i, o_{jx} \in S_j\}. \quad (3.12)$$

3. Метод дальнього сусіда:

$$D_{max}(S_i, S_j) = \max\{d(o_{ix}, o_{jx}) | o_{ix} \in S_i, o_{jx} \in S_j\}. \quad (3.13)$$

4. Центроїдний метод:

$$D_{cen}(S_i, S_j) = d(\bar{o}_i, \bar{o}_j), \quad \text{де}$$

$$\bar{o}_i = \frac{1}{N_i} \sum_{x=1}^{N_i} o_{ix} = (\frac{1}{N_i} \sum_{x=1}^{N_i} o_{ix}^{(1)}, \dots, \frac{1}{N_i} \sum_{x=1}^{N_i} o_{ix}^{(M)}), \quad (3.14)$$

$$\bar{o}_j = \frac{1}{N_j} \sum_{x=L}^{N_j} o_{jx} = \left(\frac{1}{N} \sum_{x=L}^{N_j} o_{jx}^{(1)}, \dots, \frac{1}{N} \sum_{x=L}^{N_j} o_{jx}^{(M)} \right). \quad (3.15)$$

5. Метод групового середнього (середній зв'язок):

$$D_{arg}(S_i, S_j) = \frac{1}{N_i \times N_j} \sum_{x=1}^{N_i} \sum_{y=1}^{N_j} d(o_{ix}, o_{jy}). \quad (3.16)$$

Важливим при реалізації процедур класифікації є поняття якості класифікації. Функціонал якості класифікації - це відображення множини всіх можливих систем класів на дійсну пряму.

Функціонали якості розбиття вихідної множини на класи використовуються, зокрема:

- для оцінки об'єктивної кількості класів (якщо вона заздалегідь невідома, що, як правило, і буває);
- для порівняння схем класифікацій, отриманих з використанням різних алгоритмів, і вибору найкращої;
- для безпосереднього використання в ієрархічних процедурах класифікації в якості відстаней (наприклад, в агломеративних процедурах на кожному кроці об'єднуються ті два класи, які дадуть максимум або мінімум функції якості серед усіх можливих попарних об'єднань).

Найбільш часто застосовуються такі способи розрахунку якості класифікації (при заданому числі класів K):

1. Сума попарних всередині класових відстаней:

$$Q_1 = \sum_{i=1}^K \sum_{x=1}^{N_i-1} \sum_{y=x+1}^{N_i} d(o_{ix}, o_{iy}). \quad (3.17)$$

2. Сума всередині класових квадратів відхилень об'єктів від середніх

$$Q_3 = \sum_{i=1}^K \sum_{j=1}^{N_i} d^2(o_{ij} - \bar{o}_i), \text{ де} \quad (3.18)$$

$$\bar{o}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} o_{ij} = \left(\frac{1}{N_i} \sum_{j=1}^{N_i} o_{ij}^{(1)}, \dots, \frac{1}{N_i} \sum_{j=1}^{N_i} o_{ij}^{(M)} \right). \quad (3.19)$$

Група ядерних методів націлена на виявлення згущення об'єктів в просторі ознак і раніше носила евристичний характер (поняття компактності об'єктів в просторі ознак не було формалізовано). Для ряду евристичних процедур з розвитком теорії були знайдені функціонали якості розбиття на групи і тим самим формалізовано відповідне їм поняття компактності [22]. Відповідно до цього

алгоритми класифікації, засновані на описі класів ядрами, поділяють на евристичні і оптимізаційні. Крім того, методи можна поділити за способом подачі об'єктів на вхід алгоритму. Якщо об'єкти подаються по одному (послідовно), то відповідні процедури називаються послідовними. Якщо на вхід алгоритмів подаються відразу всі об'єкти, то вони називаються паралельними. Перевагою послідовних процедур є висока швидкість роботи, паралельних - незалежність одержуваної класифікації від порядку об'єктів у вихідній кількості O .

Під ядром класу мається на увазі деякий реально існуючий або умовний найбільш «представницький» об'єкт, весь комплекс характеристик якого є еталоном даного класу. Часто алгоритми, засновані на описі класів ядрами, використовують процедуру класифікації об'єктів до ядер по мінімальності відстаней:

1. Задатися метрикою d
2. Знайти ядра класів $c_1, \dots, c_k, c_i = (c_i^{(I)}, \dots, c_i^{(M)}), i \in (1, \dots, K)$.
3. Віднести всі об'єкти до знайдених ядер c_1, \dots, c_k мінімальної відстані до них

$o_i \in S_i \leftrightarrow d(o_i, c_i) \leq d(o_i, c_j) \forall i, j \in \{1, \dots, K\}, i \in \{1, \dots, N\}$, сформував систему класів $S = \{S_1, \dots, S_K\}$.

Найбільш відомими методами ядерним методами класифікації є алгоритми к-середніх, к-медоїдів, Форель, алгоритм послідовного виділення ядер.

Ієрархічні методи класифікації націлені або на послідовне об'єднання вихідних об'єктів в заздалегідь задану або не задану меншу кількість класів, або, навпаки, на розчленування одного або декількох класів до потрібного ступеня деталізації. Процедури першого типу зуться ієрархічних агломеративних алгоритмів класифікації, другого - ієрархічних дивізімних алгоритмів класифікації. Результатом роботи ієрархічних алгоритмів класифікації є послідовність розбиттів множини O на непересічні класи $\{S^n | n = 1, \dots, I + 1\}$, де I - кількість проведених ітерацій, S^n - n -я система класів. Позначимо через K_n кількість

класів в n -й системі, $n \in 1, \dots, I + 1$. Тоді для агломеративних алгоритмів $K_n \geq K_{n+1}$, для дивізімних - $K_n \geq K_{n+1} \forall n \in \{1, \dots, I\}$.

Перевагами ієрархічних алгоритмів є можливості їх застосування без наявності апіорної інформації про властивості класів (наприклад, ядер класів або навчальних вибірок), модифікації для цілей географічного районування, застосування при невідомому числі класів і наочної візуалізації ходу і результатів класифікації на спеціальному графіку, який називається Дендрограмма.

До недоліків ієрархічних процедур слід віднести велику обчислювальну вартість їх реалізації. Даний недолік, правда, частково компенсується існуванням так званих «швидких» (або «порогових») ієрархічних алгоритмів, які в певних випадках дають ті ж результати, що і їх аналоги.

Найбільш відомими агломеративними методами є узагальнений за Колмогоровим і «швидкий» алгоритми ієрархічної класифікації. Узагальнений за Колмогоровим агломеративний алгоритм ієрархічної класифікації починає свою роботу з формування $K_i = N$ класів (при цьому кожен об'єкт на нульовому кроці представляє окремий клас) і проводить в загальному випадку $I = N - 1$ ітерацію. На кожному кроці алгоритму відбувається об'єднання двох «найближчих» класів в один, тобто $K_{n-1} = K_{n+1}$. Останній $(N - 1)$ - й крок алгоритму характеризується поєднанням двох, сформованих на попередніх етапах класів, в один клас, що включає в себе всі наявні (що надійшли на вхід аналізу) об'єкти:

$$S^n = \{S_j^n \neq \theta | j = 1, \dots, N - n + 1\}, \quad (3.20)$$

$$O = \bigcup_{i=1}^{N-n+1} S_i^n, S_i^n \cap S_j^n = \theta \forall i, j \in \{1, \dots, N - n + 1\}, i \neq j, n \in \{1, \dots, N\}.$$

Під «близькістю» двох класів може розумітися одна з відстаней, що являється окремим випадком узагальненого за Колмогоровим відстаню між класами (наприклад, відстань типу «ближнього сусіда»). Вибір відстані настільки впливає на результат класифікації, що найчастіше вона вноситься в назву алгоритму (наприклад, агломеративний алгоритм «середнього зв'язку»).

«Швидкі» агломеративні алгоритми засновані на використанні деякої заздалегідь заданої або настроєної в процесі класифікації послідовності порогових значень c_1, \dots, c_I , (при цьому цілком можливо, що $c_i = c = \text{const} \forall n \in \{1, \dots, I\}$). Найпростіший з підходів до автоматичного вибору порогових значень c_i , може полягати у використанні середньої відстані між об'єктами і номера кроку, наприклад:

$$c_n = \frac{n}{N} \sum_{i=1}^N d(o_i, o_j). \quad (3.21)$$

Дивізимні алгоритми ієрархічної класифікації зазвичай спираються на ядерні методи або використовують апарат теорії графів (методи на основі розчленування графів близькості). Так, дивізимний алгоритм ієрархічної класифікації на основі методу 2-середніх починає свою роботу з формування єдиного класу, що містить всі об'єкти, і проводить в загальному випадку $I = N-1$ ітерацію. На кожному кроці алгоритму відбувається послідовне розділення одного з класів на два таким чином, щоб якість одержуваної класифікації була максимальною. Останній етап роботи алгоритму ділити єдиний, який залишився нерозчленований клас, що складається з двох об'єктів, на два класи (по одному об'єкту в кожному).

Результатом роботи дивізимного алгоритму на основі методу 2-середніх є послідовність розбиття множини O на непересічні класи:

$$\{S^n | n = 1, \dots, N\}, \text{ де} \quad (3.22)$$

$$S^n = \{S_j^n \neq \theta | j = 1, \dots, n\}, \quad (3.23)$$

$$O = \bigcup_{i=1}^n S_i^n, S_i^n \cap S_j^n = \theta \forall i, j \in \{1, \dots, n\}, i \neq j, n \in \{1, \dots, N\}. \quad (3.24)$$

Для отримання заздалегідь заданої кількості класів K досить провести $I = K - 1$ ітерацію.

Дивізимний алгоритм розчленування графа, за аналогією з «швидким» агломеративним алгоритмом, вимагає завдання послідовності порогових значень c_1, \dots, c_I . На кожній ітерації алгоритму $n \in \{1, \dots, I\}$ відбувається виключення з графа ребер з великими значеннями, після чого перевіряється, на скільки підграфів (компонент зв'язності) розпався вихідний граф. Кожен підграф вдає із себе окремий клас.

У прикладній статистиці часто виділяється група методів класифікації при наявності обмежень на зв'язку між об'єктами. У роботі [2] наводяться різні інтерпретації матриці обмежень і модифікований алгоритм членування матриці близькості для проведення подібних класифікацій. Як буде показано нижче, в геоінформатиці цей клас методів досить поширений.

3.2. Класифікації багатовимірних даних

3.2.1 Оціночні і типологічні класифікації багатовимірних даних

У вітчизняній літературі з геоінформатики склалася певна термінологія диференціювання методів класифікації за її кінцевої мети. Зазвичай виділяють так звані оціночні і типологічні класифікації.

$$S_1 < S_2 < \dots < S_K \text{ і } S_1 > S_2 > \dots > S_K.$$

Під позначенням $S_1 < S_2$ розуміється, що будь-яка ОТО і-го класу краще ОТО j-го класу за комплексом показників. Вважається також, що в межах кожного класу ОТО приблизно однаково хороші або однаково погані (в розрізі проблеми, описуваної показниками).

Обов'язковим етапом в оціночних класифікаціях є перехід до єдиної ознаки. Результируючий показник отримують таким чином, щоб класи ОТО з його мінімальними значеннями могли інтерпретуватися як «погані», а з максимальними як «хороші», або навпаки. Наприклад, нехай в якості ОТО виступають земельні ділянки, а в якості показників - тип ґрунтів, віддаленість від водних ресурсів, близькість до транспортної мережі і т.д. Тоді цільовою ознакою може бути ступінь доцільності побудови житлового будинку. Залежно від значень цільової ознаки ОТО можуть розбиватися на три класи: «непридатні для будівництва», «середні» і «найкращі для будівництва». У роботі [176] як один з прикладів наведена оціночна класифікація регіонів Росії за ступенем погіршення екологічного стану навколишнього середовища, для класифікації використовувалося 9 показників: питома вага уловлених і знешкоджених шкідливих речовин у загальному обсязі, що відходять від стаціонарних джерел; питома вага нормативно очищених стічних вод, в загальному обсязі потребують очищення вод, що скидаються, рівень забруднення

води за хімічними і мікробіологічними показниками та ін. Результати оціночних класифікацій зазвичай використовують при прийнятті управлінських рішень або для аналізу причин тієї чи іншої ситуації в ОТО за обраними показниками.

3.2.2 Алгоритми класифікації багатовимірних даних, що застосовуються в геоінформатиці

Найбільш часто в географічній літературі зустрічаються описи дослідів проведення класифікацій за допомогою «вроцлавської таксономії» [14, 49, 93, 176]. «Вроцлавська таксономія» - один з алгоритмів розчленування графа. На основі матриці відстаней будується «дендрит» - деревовидний граф зв'язків ОТО за комплексом показників. Класи ОТО виходять шляхом розсічення найбільш довгих відстаней на дендрит.

Не менш часто географами використовуються і ієрархічні агломеративні методи класифікації [13, 49, 177, 180]. Цими методами зазвичай є класичні алгоритми далекого сусіда, ближнього сусіда, середнього зв'язку.

Ядерні методи також не обділені увагою географів. Зокрема, метод к-середніх присутній в системі методів, наведених в роботі [180]. Деякі евристичні ядерні алгоритми класифікації були запропоновані географами вже більше двох десятиріч тому [174]. У своїх роботах вони виділяють два основних евристичних принципи формування ядер: максимальна гетерогенність (відмінність) ядер - ОТО за комплексом показників або максимальна гомогенність (схожість) ОТО в межах кожного з класів. Крім того, завжди в цих роботах ядра класів відповідають реально існуючим ОТО, а алгоритми виділення ядер є паралельними евристичними.

Для реалізації першого евристичного принципу формування ядер був розроблений алгоритм, на вхід якого подається таблиця виду ОТО-ознака і обрана метрика d [180].

1. Знайти умовну ОТО $o_0 = (o_0^{(1)}, \dots, o_0^{(M)})$ середнім, комплексом показників за формулою

$$O_0^{(j)} = \frac{1}{N} \sum_{i=1}^N O_i^{(j)}. \quad (3.25)$$

2. Розрахувати матрицю ОТО-ОТО с урахуванням знайденої ОТО O_0 . Перший стовпець і перший рядок отриманої матриці відповідають відстаням від O_0 до інших ОТО.

3. Нехай вже знайдено $K-1$ ядер, $K \geq 1$. Тоді для знаходження чергового K -го ядра знайти номер j максимального елемента першого стовпця матриці відстаней A :

$$a_j = \max\{a_i | i \in 1, \dots, N + 1\}. \quad (3.26)$$

Оскільки нумерація ОТО була збита вставкою умовної ОТО, O_{j-1} покласти в якості чергового ядра.

4. Якщо необхідно отримати ще одне ядро, додати до першого стовпця матриці A j -й стовпець за формулою

$$a_{11} = a_{21} + a_{11} \quad \forall i \in \{1, \dots, N + 1\}. \quad (3.27)$$

і перейти до кроку 3.

Формально гетерогенність До ядер O_{1l}, \dots, O_{Kl} є реальними ОТО, відповідає умові

$$\begin{aligned} \underline{Q}(o_{t1}, \dots, o_{K1}) \\ &= \sum_{i=1}^{K-1} \sum_{j=i+1}^K d(o_{i1}, o_{j1}) \\ &= \max\left\{ \sum_{i=1}^{K-1} \sum_{j=i+1}^K d(o_{ti}, o_{tj}) \mid t_i \neq t_j \quad \forall i \neq j, o_t \in O \right\}. \end{aligned} \quad (3.28)$$

Для $K > 2$ описаний алгоритм обчислює гетерогенні ядра наближено. Тому для коректного знаходження гетерогенних ядер, наприклад, для $K = 4$, необхідно використовувати повний перебір всіх різних четвірок ОТО як ядер o_{t1}, \dots, o_{tK} і обчислення $\underline{Q}(o_{t1}, \dots, o_{tK})$

Можна навести приклади, коли використання максимально гетерогенних ядер буде спотворювати класифікацію (див. Додаток А). Перевагою алгоритму є висока обчислювальна ефективність і простота реалізації. Це дозволяє

використовувати його для розвідувальних класифікацій великої кількості ОТО, що необхідно для формування припущень і гіпотез про структуру простору ознак. Крім того, вибір ядер дозволяє використовувати їх в якості початкових наближень для більш складних багато ітеративних алгоритмів (наприклад, к-середніх).

Другий принцип формування ядер ґрунтується на виділенні максимально гомогенних класів ОТО. При виборі ядер класів необхідно, щоб при віднесенні до них залишилися ОТО по мінімальності відстані (і отриманні таким чином системи класів S) використовується функціонал якості класифікації Q досягав свого екстремуму (максимуму або мінімуму, в залежності від інтерпретації функціоналу якості):

$$Q(S(o_{i1}, \dots, o_{ik})) = \max\{Q(S(o_{j1}, \dots, o_{jk})) | j_1, \dots, j_k \in 1, \dots, N; j_x \neq j_y \forall x \neq y\},$$

o_{i1}, \dots, o_{ik} - шукані ядра класів,

$S(o_{i1}, \dots, o_{ik})$ - отримана відповідно до розподілу ОТО до ядер по мінімальності відстаней система класів.

Один з евристичних ядерних алгоритмів для формування гомогенних класів запропонований в роботі [179]. Алгоритм отримує на вхід матрицю A виду ОТО-ОТО, а в якості функціоналу якості використовує суму попарних міжкласові відстаней Q_2 .

1. Для знаходження перших двох ядер o_{i1}, o_{i2} знайти по матриці ОТО-ОТО дві ОТО, відстані між якими максимальні: $a_{i1,i2} = \max\{a_{ij} | j, t \in 1, \dots, N\}$

2. Шлях виділені ядра o_{i1}, \dots, o_{ik-1} $K \geq 3$. Тоді для отримання нового K -го ядра:

2.1) Випробувати кожну з «неядерних» ОТО в якості нового ядра, розподілити по мінімальності відстані інші ОТО і отримати

$N + 1$ До значень

$$Q_2(S(o_{i1}, \dots, o_{ik-1})) \forall t \in N_K = \{1, \dots, K\} | \{i_1, \dots, i_{K-1}\}, \quad (3.29)$$

Знайти $i_K \in N_K$,

$$Q_2(S(o_{i1}, \dots, o_{ik})) = \max\{Q_2(S(o_{i1}, \dots, o_{ik-1}, o_t)) | t \in N_K\}, \quad (3.30)$$

і покласти o_{i1} як K -го ядра.

3. При необхідності отримання ще одного ядра повернутися до попереднього кроку.

Як і в наведеному вище методі знаходження гетерогенних ядер, недоліком алгоритму є його не оптимальність в сенсі обраного критерію якості класифікації. Іншими словами, не завжди отримана відповідно до алгоритму система класів S звертає в максимум заданий показник якості для відомого заздалегідь кількості класів K .

Перевагою двох наведених алгоритмів є велике число їх застосувань для проведення типологічних класифікацій в географії. Зауважимо, що формально одні і ті ж методи можуть застосовуватися як для оціночних, так і для типологічних класифікацій. Для оціночних класифікацій в географії найчастіше застосовуються ієрархічні алгоритми, для типологічних - «ядерні».

3.2.3 Облік при класифікації багатовимірних даних географічного простору

Розробка методів класифікації, що дозволяють враховувати при проведенні аналізу географічний простір, ведеться в основному в руслі робіт по районуванню. Ю.Г.Саушкін визначив географічне районування як «процес виділення і вивчення об'єктивно існуючої (у природі, в розселенні і в житті людей, в господарстві) територіальної структури, організованості та ієрархічної підпорядкованості» [90]. В.І. Блануца розуміє під районуванням «процес багатфакторного (багатознакового) поділу території на множину непересічних цілісних районів, що представляють собою компактні згущення деяких вихідних осередків (точок) як в тривимірному фізичному, так і в багатовимірному просторі ознак» [12]. В.С.Тікунов розглядає районування як «процедуру виокремлення цілісних територіальних систем, коли увага дослідників концентрується на відмінностях між ними» [174].

У термінах районування синонімом класу з обмеженням на його географічну нерозчленованність є поняття району. Будемо позначати систему районів і район відповідно символами S і s , маючи на увазі, що якщо

$$S_i = \{o_{ij} \in O | j \in 1, \dots, N_i\}, \text{ тоді} \quad (3.31)$$

$$\forall o_{ij} \in S_i \exists S_t : g(o_{ij}, o_{it}) = 1, \text{ де} \quad (3.32)$$

$$j, t \in \{1, \dots, N_t\}, \quad (3.33)$$

де g - географічна відстань типу суміжності.

Хоча й існують унікальні географічні алгоритми районування (наприклад, метод бар'єрів максимальних відмінностей), найчастіше методи автоматичного районування є модифікаціями існуючих методів класифікації. У працях [180, 289] наводиться теза, що модифікація існуючих методів автоматичної класифікації для цілей районування може здійснюватися таким чином:

- шляхом зміни формул заходів близькості між об'єктами за рахунок відображення в них відносного географічного положення об'єктів, що дозволяє врахувати топологічні властивості об'єктів в заходи подібності;
- шляхом зміни формул функціоналу якості розбиття на класи, які повинні бути за змістом ширше стандартних за рахунок включення територіальної складової;
- шляхом модифікації алгоритмів автоматичної класифікації, зазвичай за рахунок включення додаткових кроків перевірки географічної суміжності об'єктів.

Найчастіше для районування користуються звичайними алгоритмами класифікації, а після отримання класів переходять до районів. Таку операцію можна проводити або вручну, або автоматичним розбиттям несуміжних частин одного класу на різні райони.

У праці Л.Я.Нутенко [100] розглядається ряд методів «регіоналізації». В англійській мові аналогом термін «районування» є термін «regionalization». Класифікація географічних об'єктів розглядається автором як завдання теорії розпізнавання образів. Найпростіші методи районування полягають у виборі деякої системи порогових значень і подальшої угруповання на основі розглянутого показника.

У праці [6] описаний варіант застосування методу інформаційних градієнтів для районування. Метод заснований на теорії інформації. Він, на відміну від методів параметричної статистики, вільний від нестачі накладення обмеження на нормальність розподілу випадкової величини. Показники аналізу повинні бути безперервно розподілені по території.

У працях [307, 312] наводиться досвід кількісної розробки схеми економічних районів. Авторами використовується розроблений ними метод випадкового пошуку. Проведене районування необхідно для правильного територіального розміщення основних груп виробничих галузей (які вимагають достатньої великій території) і визначення раціональних напрямків господарського розвитку основних типів регіонів.

Цілий спектр робіт з районування належить перу співробітників Казанського державного університету А.М.Трофімову, В.А. Рубцова і їх колегам і співавторам. Зокрема, випущені монографія [180] навчальний посібник з районування в двох частинах [180]. Автори виділяють три групи методів, на базі яких можлива реалізація методів районування [180]:

- методи математичної статистики (кореляційний, факторний аналізу та ін.);
- методи теорії розпізнавання образів (методи найближчого і далекого сусіда, k-середніх і ін.);
- методи теорії графів (методи одноярусного і багатоярусного членування).

Ряд робіт з районування, включаючи монографію [11], опублікований представником Інституту географії Сибірського відділення Російської академії наук В.І.Блануца. Так, в роботі [14] розглядається загальна схема географічної класифікації, в якій окремим етапом виділяється контроль якості. Наводяться різні заходи якості класифікацій, в яких фігурує і «простота». Хоча на терміні «простота» автор не зупиняється детально, виникає ідея застосувати його для опису розподілу об'єктів в межах класів в просторі ознак і класів в географічному просторі (на карті). Як «простоти» пропонується, наприклад, використовувати ступінь перфорування класів. Формула кількісного коефіцієнта перфорування класів присутній в роботі [306]. У разі отримання декількох угруповань

пропонується або порівнювати їх функціонали якості, або формувати на їх основі одне угруповання, для чого наводиться спеціальний алгоритм. Ставиться також завдання порівняння отриманої угруповання з «еталонної» з метою уточнення «еталона» і перевірки роботи методу. У монографії [38] наводиться модифікація ієрархічного агломеративного алгоритму для цілей районування. У працях [52,56] також наводиться розроблений автором ієрархічний агломеративний алгоритм районування, заснований на поняттях теорій графів і інформації.

У праці [73,74] описана формальна модель районування, однак районування не слід розуміти як отримання територіально нерозчленованих класів. Пропонується дотримуватися схеми детермінованого, ймовірно-статистичного та евристичного підходів. Крім того, при відсутності значення ознаки об'єкта рекомендується звертатися до аналізу його найближчих сусідів в географічному просторі.

Дослідження в області розробки методів класифікації просторових даних, природно, ведуться не тільки в нашій країні. Тут відзначимо лише найбільш цікаві результати, отримані зарубіжними вченими. Як приклад створення універсального методу класифікації закордонним географом можна привести ієрархічний метод, описаний в праці [56]. В даний час даний метод носить назву агломеративного ієрархічного алгоритму але методу групового середнього.

Параметричні методи класифікації на основі моделі суміші розподілів грають найважливішу роль в прикладній статистиці. На основі цих методів розроблені ефективні алгоритми, які можуть застосовуватися і для класифікації географічних даних. У праці [194] описана модифікація ЕМ-алгоритму, яку автори назвали NEM-алгоритмом (Neighborhood EM-алгоритм). Цей метод дозволяє враховувати, крім атрибутивного простору ознак, і будь-який інший простір, заданий матрицею близькості ОТО-ОТО. При використанні географічного простору матриця близькості може бути як бінарної таблицею суміжності, так і задавати географічні відстані між парами ОТО. Функція правдоподібності, яку потрібно максимізувати для знаходження невідомих векторів параметрів ϑ і $\bar{\vartheta}$, записується у вигляді

$$U = L(\vartheta) + \beta W(\bar{\vartheta}), \text{ де} \quad (3.34)$$

$$L(\vartheta) = \sum_{i=1}^N \ln \left(\sum_{j=1}^K \rho_i f(o_i, \vartheta_j) \right)$$

– звичайна логарифмічна функція правдоподібності,

$$W(\tilde{\vartheta}) = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \rho_k^{(i)} \rho_k^{(j)} g_{ij}, \quad (3.35)$$

β - вага (важливість) географічного простору

$G = (g_{ij})$ - матриця ОТО-ОТО географічного простору (в разі районування $g_{ij} \in \{0,1\}$,

$\rho_j^{(i)}$ - оцінки апостеріорних ймовірностей приналежності i -й ОТО j -му-класу, $i \in \{1, \dots, N\}, j \in \{1, \dots, K\}$

За аналогією з ЕМ-алгоритмом, Нем-алгоритм ітераційно знаходить оцінки всіх параметрів, на кожному кроці покращуючи їх. Оцінки апостеріорного ймовірностей $\rho_j^{(i)}$ використовуються для районування (або нечіткого районування) точно так же, як і в класичному ЕМ-алгоритмі. Недоліком алгоритму є необхідність підбору параметра ваги географічного простору β від значень якого дуже сильно залежить результат районування.

Крім стандартних методів, географами розроблені і широко використовуються спеціальні добре інтерпретуються процедури районування, для яких відсутні аналоги в класифікації. Мова в першу чергу йде про метод бар'єрів максимальних відмінностей, запропонованому Монмонье [322]. Цей метод можна застосовувати тільки до ОТО полігонального типу.

З наявністю бар'єру для заданого району (що містить більше однієї ОТО) будемо називати максимальна відстань між двома географічно суміжними ОТО цього району в просторі показників, бар'єром -географічний кордон між двома відповідними ОТО. Позначати значення бар'єру i -го району будемо символом $B(S_i)$:

$$B(S_i) = \max\{d(o_{ix}, o_{iy}) | o_{ix}, o_{iy} \in S_i, g(o_{ix}, o_{iy}) = 1\}, \quad (3.36)$$

Для аналізу даних методом бар'єрів максимальних відмінностей обчислюються відстані лише між географічно суміжними ОТО, а все ОТО при ініціалізації алгоритму відносяться до одного єдиного району. Після цього на кожному кроці відбувається поділ одного з отриманих районів на два (не обов'язково рівних) району. Починається поділ з визначення в кожному з уже отриманих районів але одного значення бар'єрів, серед яких відшукується максимальний. Район, який має в своєму складі бар'єр з максимальним значенням, підлягає поділу. По обидва боки від бар'єру ітеративно проводиться межа - до тих пір, поки одна не розділить поточний район. Чергова ітерація розширює кордон за рахунок приєднання наступного бар'єру (який обчислюється без урахування пар ОТО, вже що взяли участь в утворенні попередніх бар'єрів), що примикає до кордону.

3.3 Реалізація методів автоматичної класифікації і районування в сучасних векторних ГІС

3.3.1 Реалізація в векторних ГІС методів класифікації за єдиною ознакою

Класифікацію з розмірності простору ознак можна поділити на одновимірну ($M=1$) і багатовимірну ($M>1$). Класифікація просторових об'єктів є однією з основних функцій векторних ГІС. Проведення одновимірної класифікації необхідно для отримання тематичних карт методами «колірна шкала» для всіх типів об'єктів і «градуирований символ» для точкових і лінійних об'єктів за кількісним або рангових ознакою. Тому все векторні ГІС мають в своєму арсеналі кілька алгоритмів одновимірної класифікації. В якості таких алгоритмів найчастіше виступають методи «рівно площинна» і «рівних інтервалів». Перший метод дозволяє отримувати класи з приблизно однаковою кількістю об'єктів в кожному, другий передбачає розбиття всього діапазону зміни заданого ознаки на відрізки однакової довжини.

Як і в випадку багатовимірних класифікацій, результатом роботи методу одновимірної класифікації є розбиття множини O на непересічні класи з додатковою умовою:

При цьому для «рівно площинного» методу

$$|S_i| = N_i \quad \forall i \in \{1, \dots, K\}, \quad (3.37)$$

$$N_i \in \left\{ \left\lceil \frac{N}{K} \right\rceil, \left\lceil \frac{N}{K} \right\rceil + 1 \right\}, \text{ де} \quad (3.38)$$

$\lceil \cdot \rceil$ - операція взяття цілої частини.

Для методу «рівних інтегралів»

$$\forall o \in O \setminus O_{min} \quad \forall i \in \{1, \dots, K\} \quad o \in S_i \Leftrightarrow (i-1) \times \frac{o_{max}^{(1)} - o_{min}^{(1)}}{K} < o \leq i \times \frac{o_{max}^{(1)} - o_{min}^{(1)}}{K},$$

$$o_{min} \in S_i, \quad (3.39)$$

$$o_{min}^{(1)} = \min\{o_i^{(1)} \mid i \in 1, \dots, N\}, \quad o_{max}^{(1)} = \max\{o_i^{(1)} \mid i \in 1, \dots, N\}. \quad (3.40)$$

Крім описаних методів, деякі ГІС мають і більш витончені алгоритми, що враховують, наприклад, розкид об'єктів від їх середнього або визначають, по аналогії з багатовимірної класифікації, компактні згущення об'єктів всередині деяких відрізків-класів.

3.3.2 Реалізація в векторних ГІС методів класифікації за багатьма ознаками

Багатовимірні алгоритми класифікації в векторних ГІС майже не представлені. У ГІС ArcGIS з'явився додатковий геостатистичного модуль розширення Geostatistical Analyst і єдина функція багатовимірної класифікації. Однак інструменти Geostatistical Analyst досить специфічні, а функція класифікації орієнтована тільки на якісні ознаки і зводиться до виділення таких груп об'єктів, які мають повний збіг значень всіх ознак.

Звичайно, можливість проведення багатовимірної класифікації просторових даних (правда, без урахування їх географії) є у будь-якого дослідника. Існує достатня кількість статистичних пакетів, серед яких можна виділити SAS,

Statgraphics, SPSS та ін. З урахуванням регулярних оновлень статистичних пакетів можна бути впевненим, що більша кількість сучасних методів класифікації в них присутня. Однак технологія аналізу представляється досить трудомісткою: необхідно або розробляти в кожному конкретному випадку конвертер між ГІС і статистичним пакетом, або вручну переносити вихідні дані і результати класифікацій між ними. Крім того, врахувати при класифікації географічний простір буде в більшості випадків (наприклад, при проведенні районування) неможливо.

Інтерес користувачів математичних пакетів до можливостей просторового аналізу і картографічної візуалізації даних в ГІС, а також користувачів ГІС до функцій бізнес-аналізу був оформлений в спільну розробку компаній SAS Institute і ESRI (sas.com, support.sas.com). Відповідний програмний продукт з'явився на ринку в кінці 2002 року під назвою SAS Bridge for ESRI. Він став сполучною ланкою, що дозволяє користувачам лінійки програмних продуктів ESRI ArcGIS викликати функції математичного пакета SAS, і навпаки. Однак нових функцій просторового аналізу (наприклад, таких, як районування) в ArcGIS при цьому не з'явилося.

Найбільш серйозний крок у вирішенні проблеми використання в ГІС інструментів розвиненого статистичного аналізу був зроблений фахівцями ЩІ компаній ESRI і MathSoft. Компанія MathSoft, розробник пакета MathCAD для символьних обчислень і системи статистичного аналізу S-PLUS, випустила новий модуль для ESRI ArcView GIS - S-PLUS for ArcView GIS. Фактично це розширення забезпечує доступ з ArcView GIS до функцій пакета S-PLUS, який продовжує залишатися самостійним продуктом (dataplus.ru). В якості окремого продукту також реалізований новий модуль S + SpatialStats - додаток по просторовій статистиці для S-PLUS.

Функції представлення результатів статистичного аналізу на картографічній основі з'явилися і в статистичному пакеті SPSS. Розробники математичного програмного продукту скористалися при цьому готовими рішеннями, що поставляються фірмою MapInfo. При інсталяції пакета можна встановити і додаток SPSS Map Geoset Manager, що дозволяє маніпулювати картографічними

векторними шарами (spss.com). У стандартному постачанні, наприклад, присутні карти адміністративно-територіального поділу світу, Європи, США.

Детальний розгляд NEM-алгоритму досить громіздко і може бути знайдено у відповідній роботі [317]. Розгляд же деяких ядерних і ієрархічних алгоритмів районування необхідно для їх подальшої комп'ютерної реалізації.

3.4 Ядерні методи автоматичного районування

3.4.1 Загальна схема ядерного алгоритму районування

Основною проблемою даного класу методів є вибір ядер районів, які зобов'язані бути реальними ОТО. Після того, як ядра утворення районів отримані, починається послідовне приєднання до них максимально «схожих» суміжних ОТО. Таким чином, маємо алгоритм, схожий з наведеним раніше для випадку звичайних класифікацій.

Загальна схема алгоритму районування на основі віднесення ОТО до ядер по суміжності і мінімальності відстаней.

1. Отримати матрицю близькості ОТО-ОТО $A = (a_{ij})$ для атрибутивного простору ознак і матрицю суміжності ОТО-ОТО $G = (g_{ij})$ для географічного простору.

2. Знайти ядра класів o_{i_1}, \dots, o_{j_k} припустимо $S_i^j = \{o_{i_1}\}, \dots, S_K^1 = \{o_{i_k}\}$.

3. Припустимо $n=1$.

4. Нехай на кроці $n \in \{1, \dots, N - K\}$ отримані класи S_1^n, \dots, S_K^n . Тоді:

4.1) Знайти $j \in \{1, \dots, N\}, i \in \{1, \dots, K\}$:

$$a_{ji} = \min\{a_{ji} | \exists o_x \in S_y: g_{jx} = 1, y \in 1, \dots, K\}. \quad (3.41)$$

4.2) Припустимо $\forall i \in \{1, \dots, K\}$,

$$S_i^{n+1} = \begin{cases} S_i^n \cup \{o_j\}, & i = t \\ S_j^n, & i \neq t \end{cases} \quad (3.42)$$

4.3) Припустимо $a_{ji} = +\infty$,

5. Якщо $n = N - K$, кінець, припустимо по - іншому $n = n + 1$ і переходимо до кроку 4.

Наведений алгоритм передбачає тільки класифікацію ОТО до вже відомих, заданих ядер районів за принципом географічної суміжності і близькості в просторі ознак. Початкові ядра районів можуть знаходитися, наприклад, за допомогою експертів або описаних раніше евристичних процедур [55].

У географічних дослідженнях часто застосовують для районування стандартні методи класифікації, а після отримання класів аналізують їх просторову структуру. Потім, виділивши в кожному класі кілька ОТО, що утворюють цілісний географічний район, вважають їх ядрами районів. Після цього виділені ядра розширюються шляхом до класифікації залишилися ВТЕ за наведеним вище алгоритмом.

3.4.2 Алгоритм районування на основі методу класифікації к-методів

Можна розширити алгоритм етапом перерахунку ядер районів після закінчення роботи алгоритму і подачі на його вхід нових ядер. Таким чином, наведений алгоритм може бути частиною алгоритму районування на основі методу к-методів. Однак для такого алгоритму нема впевненості в тому, що він не зациклиться, тому що не обов'язково на наступній ітерації функціонал якості районування буде зростати. Метод к-середніх не може бути адаптований до завдань районування, тому що він оперує з гіпотетичними ОТО як центрами, для яких неможливо перевірити умови на межовість (суміжність).

Загальна схема алгоритму районування к-методів.

1. Задається метрикою d і K , початковими центрами класів $c_{11}, c_{12}, \dots, c_{1K}$ (вибравши їх, наприклад, випадковим чином, отримавши за допомогою експерта, евристичного алгоритму або обчисливши за навчальною вибіркою):

$$c_{1i} = (c_{1i}^{(1)}, \dots, c_{1i}^{(M)}), i \in \{1, \dots, K\}, \quad (3.43)$$

Відзначимо, що початкові центри класів обов'язково повинні збігатися з реальними ОТО, мати географічну прив'язку, тобто

$$c_{1i} \in O \forall i \in \{1, \dots, K\}. \quad (3.44)$$

Сформуувати матрицю суміжності $G = (g_{ij})$.

2. Покласти $n = 1$.

3. Класифікувати все ОТО до центрів c_{n1}, \dots, c_{nK} мінімальності відстані до них і суміжності в повній відповідності з наведеними вище алгоритмом.

4. Порахувати вектора середніх для кожного з класів, які не являються в загальному випадку реальними об'єктами і не мають географічної складової:

$$\bar{o}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} o_{ij}, \text{ де} \quad (3.45)$$

$$S_i = \{o_{1i}, \dots, o_{N_i i}\}, i \in \{1, \dots, K\}. \quad (3.46)$$

5. Знайти ОТО, найближчі до нових середніх по кожному класу тощо.

$$c_{n+1i} \in O: d(c_{n+1i}, \bar{o}_i) \leq d(o_{ij}, \bar{o}_i), i \in \{1, \dots, K\}, t_i, j \in \{1, \dots, N\}. \quad (3.47)$$

У разі досягнення мінімуму відразу на двох і більше ОТО, включити рандомізацію.

6. Якщо $c_{ni} = c_{n+1i} \forall i \in \{1, \dots, K\}$, кінець, припустити по - іншому $n = n + 1$ і перейти до кроку 3.

У разі, якщо число ОТО і ознак велике, кількість ітерацій обмежують або заздалегідь заданим числом, або відсотками зміни знову отриманих на кроці 3 метоїдів від попередніх. Так, наприклад, це зроблено в пакеті ERDAS IMAGINE.

3.4.3 Ізоморфізм постановок задач класифікації багатовимірних даних при наявності обмежень

Повертаючись до прикладів не географічних постановок задач районування, відзначимо їх подібні смислові риси. Ядра районів в термінології завдання формування експертних груп - це деякий лідер, який має найбільш типові професійні характеристики і хороші взаємини з більшістю інших експертів. Відповідно, завданню класифікації гіпертекстів ядром класу може виступати

документ з максимальною кількістю посилань на інші документи (тобто об'єкт, який максимально межує з іншими можливими представниками класу) і має типові значення ознак в атрибутивному просторі. Крім того, в географії одним з принципів районування є отримання районів приблизно однакової площі (тобто з приблизно однаковою кількістю ОТО в кожному). У задачі формування експертних груп одержувані комісії також повинні бути співмірними.

До перспективних напрямків розвитку «ядерних» методів районування слід віднести розробку алгоритмів автоматичного отримання ядер класів, виходячи з конфігурації простору ознак і географічного простору. В даний час при проведенні практичних досліджень ядра зазвичай отримують експертним методом або за допомогою одного з універсальних «ядерних» методів (наприклад, k - медоїдів), що враховують тільки простору ознак.

3.5 Ієрархічні методи автоматичного районування

3.5.1 Агломеративні ієрархічні алгоритми районування

Відмінність агломеративних алгоритмів районування від відповідних алгоритмів класифікації полягає в формуванні матриці географічної суміжності і перевірці на кожному кроці межовості об'єднуваних районів. При цьому обидва агломеративні алгоритми (узагальнений за Колмогорова і швидкий) можна використовувати для районування за умови, що мінімум відстані між районами (або максимум функції якості при об'єднанні) шукається тільки для географічно суміжних районів.

Схема роботи узагальненого за Колмогоровим агломеративного алгоритму районування.

1. Задається метрикою d і стратегією об'єднання D .
2. Сформувані першу систему S_1 з N районів:

$$S^1 = \{S_i^1, \dots, S_N^1\}, S_i^1 = \{o_i\}, \dots, S_N^1 = \{o_N\}, (3.48)$$

Верхній індекс в позначенні районів буде вказувати крок алгоритму, нижній - номер району в поточній системі. Припустимо $n = 1$.

3. Припустити $K = N - n + 1$ і отримати матрицю суміжності $G_{K \times K}^n$:

$$g^n(i, j) = \begin{cases} 1, \exists o_{ix} \in S_i^n, \exists o_{iy} \in S_j^n: g(o_{ix}, o_{iy}) = 1, \\ 0, \text{ або} \end{cases} \quad (3.49)$$

$$i, j \in \{1, \dots, K\}$$

4. Нехай на кроці $n \in \{1, \dots, N - 1\}$ отримана система районів

Тоді:

4.1) Обчислити відстань між районами $D(S_i^n, S_j^n) \forall i, j \in \{1, \dots, K\}$.

4.2) Знайти $x \neq y, x, y \in \{1, \dots, K\}$:

$$D(S_x^n, S_y^n) = \min\{D(S_i^n, S_j^n) | i \neq j, g^n(i, j) = 1\}. \quad (3.50)$$

4.3) Без обмеження спільності» $< j$. Припустимо

$$S_i^{n+1} = \begin{cases} S_i^n, & i \in \{1, \dots, x - 1\} \cup \{x + 1, \dots, y - 1\}, \\ S_x^n \cup S_y^n, & i = x, \\ S_{i=1}^n, & i \in \{y + 1, \dots, K\} \end{cases} \quad (3.51)$$

сформував, таким чином, нову систему районів

$$S^{n+1} = \{S_1^{n+1}, \dots, S_{K-1}^{n+1}\}. \quad (3.52)$$

5. Якщо $n + 1 = N$, кінець, інакше припустимо $n = n + 1$ и перейти к шагу 3.

Цілком може статися, що для $n < N$ на кроці 4.2 не вдасться знайти два суміжних класи, які потрібно об'єднати. Ця ситуація відповідає випадку незв'язності графа G географічних відстаней, побудованого на основі матриці суміжності G_{N+N}^1 . Якщо c - число компонент зв'язності графа G , то критерієм зупинки алгоритму на 5-му кроці є порівняння n з числом $(N - c + 1)$.

Схему швидкого агломеративного алгоритму районування тут наводити не будемо. Він також повністю відповідає швидкому агломеративного алгоритму класифікації з додаванням етапу перевірки на кроці n межовості об'єднуються районів, відстані між якими в просторі ознак менше константи c_n .

3.5.2. Дивізімні ієрархічні алгоритми районування на основі класичних дивізімних методів класифікації

Більшість дивізімних алгоритмів районування, по аналогії з «ядерними» і агломеративного, виходять шляхом природної модифікації класичних методів класифікації.

Дивізімному алгоритму класифікації, заснованого на методі 2-середніх, відповідає повністю повторює його метод районування на основі алгоритму 2-метоїдів (тому що метод 2-середніх, на відміну від 2-метоїдів, не дозволяє перейти до районів). Алгоритм районування на основі методу 2-метоїдів дозволяє побудувати розбиття одного класу на два.

Загальна схема роботи дивізімного ієрархічного алгоритму районування на основі методу районування 2-метоїдів.

1. Ініціалізувати початкові значення алгоритму.

1.1) задається метрикою d і функціоналом якості Q .

1.2) Сформувати систему районів $S^1 = \{S_i^1\}, S_i^1 = \{o_1, \dots, o_N\}$.

Верхній індекс в позначенні системи районів буде вказувати тег алгоритму, нижній - номер району в системі.

1.3) Припустимо $n = 1$.

2. Нехай на кроці $n \in \{1, \dots, N-1\}$ отримана система районів $S^n = \{S_1^n, \dots, S_n^n\}$.

Тоді:

2.1) $\forall i \in \{1, \dots, n\}$ допомогою методу районування 2-метоїдів отримати розбиття $S_1^n = S_{i1}^n \cup S_{i2}^n$.

2.2) $\forall i \in \{1, \dots, n\}$ обчислити

$Q(S^{ni})$, де

$$S^{ni} = \{S_1^n, \dots, S_{i-1}^n, S_{i1}^n, S_{i2}^n, S_{i+1}^n, \dots, S_n^n\} \quad (3.53)$$

Без обмеження спільності великі значення 0 відповідають кращій якості районування. Знайти $j: Q(S^{nj}) = \max\{Q(S^{ni}) | i = 1, \dots, n\}$.

2.3) Припустимо

$$S_i^{n+1} = \begin{cases} S_i^n, i \in \{1, \dots, j\}; \\ S_{j1}^n, i = j; \\ S_{j2}^n, i = j + 1; \\ S_{i-1}^n, i \in \{j + 2, \dots, n + 1\} \end{cases} \quad (3.54)$$

Сформулюємо таким чином нову систему районів.

$$S^{n+1} = \{S_1^{n+1}, \dots, S_{n+1}^{n+1}\} = \{S_1^n, \dots, S_{j-1}^n, S_{j1}^n, S_{j2}^n, S_{j+1}^n, \dots, S_n^n\}. \quad (3.55)$$

2. Якщо $n = N - 1$, кінець, інакше припустимо $n = n + 1$ і перейти до кроку 2.

Як видно з алгоритму, на кожній ітерації на кроці 2.1 необхідно обчислювати розбиття тільки для знову отриманих районів (інші вже існують). Замість поділу класів на дві частини за допомогою методу 2-метоїдів можна скористатися будь-яким іншим методом районування.

Дивізімному алгоритму, заснованого на розчленуванні графа, відповідає давно описаний в стандартній літературі з прикладної статистики [3] метод класифікації при обмеженнях. Обмеження накладаються на використання в дивізімном алгоритмі матрицю близькості A простору ознак типу ОТО-ОТО матрицею географічної суміжності Z відповідно до формули

$$a_{ij} = \begin{cases} +\infty, g_{ij} = 0 \\ a_{ij}, g_{ij} = 1 \end{cases} \quad (3.56)$$

Ця умова відповідає видаленню з графа відстаней признакового атрибутивного простору (відповідного матриці A) ребер, що з'єднують географічно несуміжні ОТО. Після видалення ребер до отриманої матриці застосовують класичний дивізімний алгоритм на основі розчленування графа без жодних модифікацій.

3.5.3 Дивізімний ієрархічний алгоритм районування бар'єрів максимальних відмінностей

Крім стандартних методів, географами розроблені і широко використовуються спеціальні добре інтерпретуються процедури районування, для яких відсутні аналоги в класифікації. Мова в першу чергу йде про метод бар'єрів максимальних відмінностей, запропонованому Монмонье [322]. Цей метод можна

застосовувати тільки до ОТО полігонального типу. Вище було дано визначення бар'єру і значення бар'єру для заданого району, яке позначалося як

$$B(S_1) = \max(d(0_{1x}, 0_{1y}) | 0_{1x}, 0_{1y} \in S_1, g(0_{1x}, 0_{1y}) = 1). \quad (3.57)$$

Нехай $S = \{0_1, \dots, 0_N\}$ - район, який підлягає поділу на райони

$S_1 = \{0_1, \dots, 0_{N_1}\}$ и $S_2 = \{0_{N_1+1}, \dots, 0_N\}$, $N_1 + N_2 = N$. Нижче подана схема метода бар'єрів максимальних відмінностей для поділу району на 2 частини

1. Сформувати матрицю географічної суміжності G_{N+N} , сформувати матрицю близькості атрибутивного простору ознак $A = (d_g)$

2. Знайти бар'єр району S

$$B(S) = d(0_x, 0_y) = \max\{d(0_x, 0_y) | 0_x, 0_y \in S, g(0_x, 0_y) = 1\}. \quad (3.58)$$

3. Припустимо $S_1 = \{0_x\}, S_2 = \{0_y\}$

4. Сформувати (перетворити) множину S всіх ОТО району суміжних одночасно з кожним з знову утворених районів S_1 , и S_2 :

$$\mathfrak{S} = \{0_i \in S | E \in S_1 : g(0_i, 0_E) = 1 \text{ и } E \in S_2 : g(0_i, 0_E) = 1\}. \quad (3.59)$$

Отримати черговий максимальний бар'єр, що примикає до вже знайденим, тобто $0 \in \mathfrak{S}, 0_i \in S_1, j \in \{1, 2\}$.

$$d(0_v, 0_j) = \max\{d(0_j, 0_q) | 0_j \in \mathfrak{S}, 0_q \in S_1 \cup S_2, g(0_j, 0_q) = 1\} \quad (3.60)$$

6. Припустимо $S = S / \{0_v\}$ Если $j = 1$, припустимо $S_2 = S_2 \cup \{0_v\}$. Якщо $j = 2$, припустимо $S_1 = S_1 \cup \{0_v\}$.

7. Перейти до кроку 4.

8. Віднести останні ОТО до районів S_1 , або S_2 за принципом суміжності, тобто

8.1) Сформувати множину ОТО з S , суміжних з районами S_1 , або S_2 :

$$\Omega_1 = \{0_1 \in S | E 0_1 \in S_1, g(0_1, 0_E) = 1\}, \quad (3.61)$$

$$\Omega_2 = \{0_2 \in S | E 0_2 \in S_2, g(0_2, 0_E) = 1\}. \quad (3.62)$$

8.2) Якщо $\Omega_1 = \Omega_2 = \emptyset$, завершити роботу алгоритму.

8.3) Припустимо

$$S_1 = S_1 \cup \Omega_1, S_2 = S_2 \cup \Omega_2, S = S / (\Omega_1 \cup \Omega_2). \quad (3.63)$$

і перейти до кроку 8.1.

З урахуванням наведеного алгоритму можна виписати загальну схему роботи дивізімного алгоритму районування на основі методу бар'єрів максимальних відмінностей.

1. Задається метрикою d .

2. Сформувані першу систему районів $S^1 = \{S_j^1\}, S_j^1 = \{0_j, \dots, 0_N\}$. Верхній індекс в позначенні району буде вказувати крок алгоритму, другий-номер району. Сформувані матрицю суміжності $G_{N \times N}$, припустимо $n = 1$.

3. Нехай на кроці $n \in \{1, \dots, N - 1\}$ отримана система районів $S^n = \{S_1^n, \dots, S_v^n\}$ тоді:

3.1) Обчислити $B(S_1^n) \forall i \in \{1, \dots, n\}$.

3.2) Знайти $B(S_1^n) = \max\{B(S_1^n) | i \in 1, \dots, n, j \in \{1, \dots, n\}\}$.

Таким чином, u - номер району, який буде розділений на два більш дрібних.

3.3) Провести кордон максимальних відмінностей між ОТО u -го класу, отримавши таким чином два нових райони S_j^{n+1} і S_{j+1}^{n+1} .

4 Припустимо

$$S_j^{n+1} = \begin{cases} S_j^n, i \in \{1, \dots, j-1\} \\ S_{j-1}^n, i \in \{j+2, n+1\} \end{cases} \quad (3.64)$$

сформувавши, таким чином, нову систему районів:

$$S^{n+1} = \{S_1^{n+1}, \dots, S_{n+1}^{n+1}\}. \quad (3.65)$$

5. Припустимо $n = n + 1$. Якщо $n = N$, кінець, інакше до кроку 3.

Можна обмежити кількість ітерацій алгоритму, або заздалегідь задаючи шукане число класів K , або вводячи деякий поріг c . Оскільки значення бар'єру максимальних відмінностей зменшуються з кожним кроком алгоритму, критерієм зупинки алгоритму є умова

$$B(S_1^n) = \max\{B(S_1^n) | i \in 1, \dots, n\} < c \quad j \in \{1, \dots, n\}. \quad (3.66)$$

3.6. Отримання, зберігання та використання матриці суміжності

Практично будь-яка ГІС має в своєму арсеналі інструмент просторових запитів, за допомогою якого для кожного просторового об'єкта можна визначити список географічно межуючих з ним об'єктів і задати тим самим матрицю суміжності

$$G_{N+N} = (g_{1g}),$$

$$g_{j1} = g_{1g} = g(0_1, 0_j) = \begin{cases} 1, o, u, o, \text{ суміжні} \\ 0, \text{ інакше} \end{cases} \quad (3.67)$$

Для об'єкта точкового типу суміжними з ним рахуються всі об'єкти даного шару, віддалені від нього менш ніж на задане граничне значення

$$0_1 = (x_1, y_1), 0_j = (x_j, y_j), i, j \in \{1, \dots, N\}, \quad (3.68)$$

$$g_{1j} = \begin{cases} 1, \sqrt{(x_j, y_j) + (y_1, y_1)} < E \\ 0, \sqrt{(x_j, y_j) + (y_1, y_1)} < E \end{cases}. \quad (3.69)$$

Для об'єкта полігонального типу суміжними з ним рахуються всі об'єкти даного шару, які мають з обраним об'єктом спільний кордон.

$$g_{ij} = \begin{cases} 1, o, u, o, \text{ мають спільну границю} \\ 0, \text{ інакше} \end{cases}. \quad (3.70)$$

При програмній реалізації модуля розрахунку матриці суміжності важливо врахувати той факт, що один досліджуваний об'єкт може бути представлений на векторному шарі декількома картографічними об'єктами. Ця негативна властивість просторових об'єктів може нівелюватися за рахунок використання унікальних ідентифікаторів.

Виділимо наступні проблеми реалізації системи автоматичного районування, пов'язані з матрицею суміжності.

1. Отримання матриці суміжності. Матрицю суміжності можна:

- розраховувати безпосередньо перед початком проведення кожного районування;
- розраховувати, редагувати і зберігати одного разу для кожного шару карти, використовуючи потім її при проведенні районування.

Кожен з цих підходів має свої переваги і недоліки, які будуть розглянуті нижче.

2. Коригування матриці суміжності для деяких типів карт. Найбільш часто в таке корегування потребують карти адміністративно-територіального поділу, об'єкти яких представляють країни, регіони або області. Наприклад, деякі країни або регіони можуть бути сусідами через водний транспортний шлях (тільки по морю або через річку). Так, для карти адміністративно-територіального поділу Росії існують два об'єкти (Сахалінська і Калінінградська області), які не межують ні з одним іншим об'єктом. Без вирішення проблеми коригування матриці суміжності при проведенні автоматичного районування вони будуть віднесені до відповідних їм райони з єдиним представником.

3. Адаптація матриці суміжності до таблиці аналізу «об'єкт-ознака» або «об'єкт-об'єкт». Для використання в алгоритмі районування матриці суміжності спільно з таблицею аналізу необхідно погоджувати порядок проходження їх об'єктів. Один і той же шар карти може використовуватися спільно з різними таблицями аналізу, в той час як кожна таблиця аналізу пов'язана тільки з один шаром векторної карти (зв'язок один- до багатьох).

Кожен з двох підходів до отримання матриці суміжності має плюси і мінуси їх використання. Порівняння зазначених підходів корисно узагальнити наступною таблицею.

Таблиця 3.2

Порівняння підходів до отримання матриці суміжності

Підхід	Розрахунок матриці суміжності безпосередньо перед	Попередній одноразовий розрахунок матриці
--------	---	---

	початком роботи алгоритму	суміжності для кожного шару карти
Плюси	1. Пропуск етапу адаптації матриці суміжності до таблиці аналізу 2. Відсутність необхідності зберігання розрахованої матриці суміжності	1. Скорочення часу проведення кожної класифікації за рахунок використання вже розрахованих матриць суміжності 2. Можливість ручного редагування розрахованої і збереженою матриці суміжності
Мінуси	Збільшення часу проведення кожної класифікації за рахунок попереднього етапу отримання матриці суміжності 2. Неможливість внесення ручних змін в матрицю суміжності	1. Необхідність адаптації матриці суміжності до кожної нової таблиці аналізу 2. Складність зберігання матриці суміжності для великого (більше 255) кількості об'єктів.

Найбільш часта операція, виконувана дослідником при районуванні - проведення аналізу з заданими параметрами. Для одних і тих же шару карти і таблиці аналізу дослідник може варіювати набір ознак аналізу (виділяючи їх підмножина), алгоритм районування, метрику, спосіб нормування і ваги ознак, параметри проведення методу головних компонент і т.д. Тому основним показником для вирішення питання про спосіб отримання матриці суміжності при

проведенні автоматичного районування вважатимемо час проведення районування з заданими параметрами.

3.7 Методи багатовимірної кластеризації даних

Внаслідок багатовимірності даних, що описують об'єкти дослідження, в даному класі задач моделювання однорідних територіальних зон для використання геоінформаційних систем потребується застосування додаткових методів аналізу багатовимірних сукупностей даних в рамках дисертаційного дослідження. Перспективними розглядаються методи багатовимірної кластеризації даних, які на відміну від методів статистичного аналізу не висувають ніяких вимог обсягу аналізованої вибірки, а значить, можуть бути застосовні в умовах малого обсягу даних.

Для аналізу багатовимірних експериментальних, даних традиційно застосовуються різні статистичні методи, такі як багатовимірне шкалювання [77], факторний і компонентний [50], аналіз, кореляційний [78] аналіз, байєсовський [73-75] статистичний метод. Використання цих методів дозволяє виявити цікаві для дослідника статистичні закономірності даних і відобразити ці закономірності на цифровій карті [178]. Однак в разі малого об'єму вибірки досліджуваних даних застосування цих статистичних методів не представляється можливим. У зв'язку з цим-перспективним розглядається застосування методів багатовимірної автоматичної кластеризації даних [76-80], які дозволяють розділити всі об'єкти дослідження на однорідні за своїми властивостями класи. Основна перевага методів автоматичної кластеризації перед статистичними методами аналізу даних полягає, в тому, що вони не пред'являють особливих вимог [81, 82] до обсягу або закону розподілу вихідних даних, які не залежать від ступеня корельованості характеристик об'єктів дослідження.

В даний час в літературі широко використовуються два терміни - кластеризація і класифікація. Згідно [281] під класифікацією розуміється процедура віднесення кожного з об'єктів дослідження до одного або кількох із заздалегідь визначених класів. Таку процедуру часто [100] називають

класифікацією з навчанням. Під кластеризацією [60] розуміється процедура поділу багатьох об'єктів, які досліджуються на підмножини (класи, кластери) або, іншими словами, віднесення кожного з об'єктів дослідження до одного або декількох класів (кластерів), число яких заздалегідь не визначено.

Згідно [199], синонімами терміна кластеризація є автоматична класифікація, таксономія, розпізнавання без навчання та ін. Під багатовимірною кластеризацією будемо розуміти [61,62] процедуру поділу всієї множини об'єктів на класи, здійснювану за сукупністю атрибутів (непросторових характеристик) об'єкта дослідження в багатовимірному просторі ознак. При цьому [66] в один клас об'єднуються об'єкти, подібні між собою в певному сенсі. Ступінь подібності об'єктів, що належать до одного класу, повинна бути більше, ніж ступінь подібності об'єктів, що відносяться до різних класів.

Згідно [79, 85] визначення подібності (однорідності) об'єктів є, одним з найбільш важких і найменш формалізованих етапів у вирішенні завдання кластеризації, від якого буде залежати остаточний варіант розбиття об'єктів на класи (з чого складаються класи) при даному алгоритмі розбиття. В загальному випадку, поняття однорідності об'єктів [85] задається деякою функцією, що характеризує ступінь, близькості, об'єктів або відстань між об'єктами. Функцією відстані між об'єктами і X_i називається [79, 84, 86] невід'ємна істотнознавча функція $d(X_i, X_j)$, якщо виконуються наступні умови:

- 1) $d(X_i, X_j) \geq 0$ для будь-яких X_i і X_j ;
- 2) $d(X_i, X_j) = 0$ тоді і тільки тоді, коли $X_i = X_j$;
- 3) $d(X_i, X_j) = d(X_j, X_i)$;
- 4) $d(X_i, X_j) \leq d(X_i, X_k) + d(X_k, X_j)$, де X_k , X_j , X_i будь-які три об'єкти кластеризації.

Для кластеризації об'єктів можуть використовуватися різні функції відстані (міри відстані, міри близькості). Розглянемо найбільш часто використовувані з них [215, 228, 246]:

- евклідова відстань є геометричним відстанню в багатовимірному просторі ознак об'єкта і обчислюється за формулою:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}, \quad (3.71)$$

де i, j - порядковий номер об'єкта; k - номер ознаки об'єкта; m - кількість ознак об'єкта;

- зважена евклідова відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^m \omega_k (x_{ik} - x_{jk})^2}, \quad (3.72)$$

де ω_k - вага k -ї ознаки об'єкта $\sum_{k=1}^m \omega_k = 1$,

- манхеттенська відстань або відстань міських кварталів є різницею за координатами в просторі ознак об'єкта:

$$d_U = \sum_{k=1}^m |x_{ik} - x_{jk}|, \quad (3.73)$$

- відстань Махаланобіса розраховується за формулою:

$$d_{ij} = \sqrt{(X_i - X_j)^T \Sigma^{-1} (X_i - X_j)}. \quad (3.74)$$

де X_i - вектор стовпець всіх ознак i -го об'єкта; Σ - коваріаційна матриця;

- супремум-норма розраховується за формулою:

$$d_{ij} = \sup |x_{ik} - x_{jk}|, \quad (3.75)$$

Відстань Махаланобіса (1.4), як зазначено в [81], має деякі особливості, які необхідно враховувати при його використанні. А саме, наявність у формулі матриці коваріацій, яка робить відстань між двома точками залежним від відстані між іншими точками, тому наявність аномальних значень ознаки може спотворити всю матрицю відстаней [261, 266].

Існують інші заходи близькості [298], які не є відстанями з точки зору умов 1) - 4), які, проте, широко застосовуються на практиці. Надалі будемо їх також називати мірою відстані, як це робиться в [299]. Розглянемо деякі з них:

- квадрат евклідова відстані обчислюється за формулою:

$$d_y^2 = \sum_{k=1}^m (x_{ik} - x_{jk})^2, \quad (3.76)$$

- міра Джеффріса-Матусіті розраховується за формулою:

$$d_{ij} = \sqrt{\sum_{k=1}^m (\sqrt{x_{ik}} - \sqrt{x_{jk}})^2}. \quad (3.77)$$

У тих випадках, коли необхідно провести кластеризацію об'єктів, ознаки яких виміряні в різних одиницях виміру (кг, м, Па, та ін.), То для коректного використання заходів відстані потрібне проведення нормування [330] показників, яка дозволяє перевести їх (значення показників) в безрозмірні величини. Нормування є перехід до деякого однакового опису для всіх ознак шляхом введення нової умовної безрозмірної одиниці виміру. Наведемо деякі найбільш поширені [376] способи нормування показників:

$$z_i = \frac{(x_i - \bar{x})}{\sigma} \quad , \quad (3.78)$$

$$z_i = \frac{x_i}{\bar{x}} \quad , \quad (3.79)$$

$$z_i = \frac{x_i}{x_{max}} \quad , \quad (3.80)$$

$$z_i = \frac{x_i - \bar{x}}{x_{max} - x_{min}} \quad . \quad (3.81)$$

де z_i - нове нормоване значення показника, x_i - початкове значення показника, \bar{x} - середнє значення показника по всіх зразкам, σ - середньоквадратичне відхилення значень показника по всіх зразкам, x_{max} - найбільше значення показника, x_{min} - найменше значення показника.

В даний час розроблено велику кількість алгоритмів кластеризації [392-400], що реалізують ієрархічні і неієрархічні (послідовні, паралельні і ін.) Методи. Ієрархічні алгоритми, згідно [411], є найбільш часто використовуваними і реалізуються у вигляді такої послідовності дій.

1. На основі обраної міри близькості розраховується вихідна матриця відстаней [333, 346] виду:

$$\begin{vmatrix} 0 & d_{12} & d_{13} & \dots & d_{1N} \\ 0 & 0 & d_{23} & \dots & d_{2N} \\ 0 & 0 & 0 & \dots & d_{3N} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 \end{vmatrix} \quad (3.82)$$

2. По матриці відстаней визначається мінімальна відстань d_{ij} між двома класами p і q
3. Класи p і q об'єднуються в один загальний клас r .
4. У матриці відстаней викреслюється рядок і стовпчик з номерами i та j відповідно.
5. Складається нова матриця відстаней, в якій класи p і q замінюються новим класом r .
6. Відстань d_{sr} , від будь-якого збереженого класу s до нового класу r перераховується формулою Ланса-Вільямса [99-100]:

$$d_{sr} = \alpha_p * d_{ps} + \beta * d_{pq} + \gamma |d_{ps} - d_{qs}|, \quad (3.83)$$

де d_{ps} – відстань між класами p і s ; d_{qs} – відстань між класами q і s ; d_{pq} – відстань між класами p і q ; $\alpha_p, \alpha_q, \beta, \gamma$ - коефіцієнти, що визначають метод розрахунку нових відстаней, чисельні значення яких для різних алгоритмів наведені нижче.

7. Пункти п.2-п.6 повторюються до тих пір, поки всі об'єкти не будуть розділені на задану кількість класів.

При кластеризації з використанням ієрархічних алгоритмів кожен об'єкт розглядається як одноелементний клас, при цьому правило об'єднання класів для кожного алгоритму відрізняється.

Розглянемо кілька найбільш поширених алгоритмів. В [177, 226, 392] описується алгоритм максимально локальної відстані, званий також алгоритмом найближчого сусіда, в якому класи послідовно групуються по наступним правилом: два клас об'єднуються, якщо максимальна відстань між точками одного класу і точками іншого мінімальна. Для цього алгоритму коефіцієнти у формулі (3.84) приймають наступні значення:

$$\alpha_p = \alpha_q = 0.5, \beta = 0, \gamma = 0.5. \quad (3.84)$$

В алгоритмі мінімальної локальної відстані [76, 81, 84, 87] або далекого сусіда два класи групуються, якщо найменша відстань між точками одного класу і точками іншого мінімальна. Для цього алгоритму коефіцієнти у формулі (3.85) приймають наступні значення:

$$\alpha_p = \alpha_q = 0.5, \beta = 0, \gamma = -0.5 . \quad (3.85)$$

В [81,84,86,87] описаний алгоритм групової середньої відстані, званий також алгоритмом середнього зв'язку, де об'єднуються такі два класи, у яких середнє значення відстані між усіма парами об'єктів мінімальне. Для цього алгоритму коефіцієнти у формулі (3.86) приймають наступні значення:

$$\alpha_p = \frac{n_p}{n_p+n_q}, \alpha_q = \frac{n_q}{n_p+n_q}, \beta = \gamma = 0 , \quad (3.86)$$

де n_p, n_q — число об'єктів в класах p і q (3.86) відповідно.

В [362, 382] описується алгоритм простої середньої відстані, який аналогічний алгоритму групової середньої відстані, за винятком того, що тут при об'єднанні двох класів передбачається, що розміри цих класів рівні, тобто де $n_p = n_q$. Коефіцієнти у формулі (3.87) для цього алгоритму приймають наступні значення:

$$\alpha_p = \alpha_q = 0.5, \beta = \gamma = 0 . \quad (3.87)$$

В [81,84,86] описується алгоритм уентроїдної відстані, в якому об'єднуються класи, які мають мінімальну відстань між центрами. Для цього алгоритму коефіцієнти у формулі (3.88) приймають наступні значення:

$$\alpha_p = \frac{n_p}{n_p+n_q}, \alpha_q = \frac{n_s+n_q}{n_s+n_p+n_q}, \beta = \frac{n_s}{n_s+n_p+n_q}, \gamma = 0 , \quad (3.88)$$

де n_p, n_q - число об'єктів в класах p і q відповідно.

В алгоритмі медіанної відстані [81, 84, 87] або медіанного зв'язку, групування класів проводиться аналогічно алгоритму центроїдної відстані, за винятком того, що тут при об'єднанні двох класів передбачається, що розміри цих класів рівні, тобто $n_p = n_q$.

Коефіцієнти у формулі (3.89) для цього алгоритму рівні:

$$\alpha_p = \alpha_q = 0.5, \beta = -0.25, \gamma = 0 . \quad (3.89)$$

В [79, 81, 86] описується алгоритм відстані Уорда у якому на кожному кроці групуються такі два класи, об'єднання яких призводить до мінімального збільшення суми квадратів відстаней між кожною точкою - в класі і середньою

точкою по класу. Для цього алгоритму коефіцієнти у формулі приймають наступні значення:

$$\alpha_p = \frac{n_s + n_p}{n_s + n_p + n_q}, \alpha_q = \frac{n_q}{n_p + n_q}, \beta = \frac{n_p * n_q}{(n_p + n_q)^2}, \gamma = 0, \quad (3.90)$$

де n_p, n_q - число об'єктів в класах p і q відповідно.

Алгоритми багатовимірної ієрархічної кластеризації мають необхідність обчислення і зберігання матриці відстаней протягом всієї процедури кластеризації. Так, наприклад, при кластеризації 500 об'єктів потребується зберігання і неодноразовий перегляд матриці, що містить близько 125000 елементів. Іншим недоліком ієрархічних алгоритмів є те, що в них об'єкти розподіляються за класами лише за один прохід, а тому "погане" початкове розбиття множини об'єктів не може бути покращено на наступних кроках процесу кластеризації.

Розглянемо далі найбільш часто використовувані неієрархічні алгоритми кластеризації, до яких відносяться алгоритми б-середніх і гіперсферу. Алгоритм k -середніх, який реалізує ітеративний метод кластеризації, виконується у вигляді такої послідовності дій [224, 316].

1) Вибирається k найбільш віддалених одна від одної точок, які призначаються початковими класами.

2) Значення центрів і ваг початкових класів визначається за формулами:

$$e_k^0 = x_k,$$

$$W_k^0 = 1.$$

3) Вибирається чергова точка x_i яка приєднується до класу з найближчим центром, на основі обраної міри близькості.

4) Центр нового класу перераховується по формулі:

$$e_k^j = \frac{w_k^{j-1} * e_k^{j-1} + x_i}{w_k^{j-1} + 1}. \quad (3.91)$$

5) Вага нового класу збільшується на 1:

$$w_k^j = w_k^{j-1} + 1. \quad (3.92)$$

6) Пункти п.3-п.5 повторюються до тих пір, поки всі об'єкти не будуть розділені на класи.

7) Проводиться порівняння центрів класів початкових і кінцевих:

$$e_k^m - e_k^{m-1} < \gamma. \quad (3.93)$$

8) Якщо різниця початкових і кінцевих центрів класів не перевищує деякого порогового значення γ , то кінець алгоритму, інакше значення початкових центрів і ваг класів оновлюються и пункти п.3-п.8 повторюються:

$$e_k^{m+1} = e_k^m, \quad (3.94)$$

$$w_k^{m+1} = 1. \quad (3.95)$$

Алгоритм гіперсфер [81,87], званий також алгоритмом пошуку згущень, реалізується у вигляді такої послідовності дій:

1) На основі обраної міри близькості розраховується матриця відстаней [356, 382]

2) Задається радіус гіперсфер для всіх точок X_i :

$$\rho_i = \max_j d(X_i, X_j) - \delta. \quad (3.96)$$

де $d(X_i, X_j)$ - відстань між об'єктами X_i і X_j - деяка порогова величина, яка впливає на кількість класів.

3) Для всіх точок X , будується гіперсфера радіусу ρ_i .

4) Вибирається така гіперсфера, всередині якої розташована найбільша кількість об'єктів.

5) Об'єкти, що потрапили всередину гіперсфери, оголошуються новим класом і виключаються з подальшого розгляду.

6) Пункти п.4-п.6 повторюються до тих пір, поки всі об'єкти не будуть розподілені по класам.

Алгоритми неієрархічної кластеризації також володіють деякими недоліками. Одним з основних недоліків є чутливість до викидів - одиничних високих значень будь-якої характеристики об'єкта.

Розглянемо більш докладно етапи, передбачені в концептуальній схемі побудови математичної моделі однорідних територіальних зон. На першому етапі

проводиться автоматичне розбиття (кластеризація) всієї множини об'єктів дослідження. Кластеризація об'єктів здійснюється в багатовимірному просторі ознак з використанням тільки атрибутивних (непросторових) характеристик. Це дозволяє розділити всі об'єкти на непересічні класи, таким чином, що кожен клас буде складатися тільки з однорідних (схожих) за своїми непросторовими властивостями об'єктів. При цьому об'єкти, що входять в різні класи, будуть мати великі відмінності, ніж об'єкти, що становлять один клас. За результатами кластеризації для кожного об'єкта дослідження вводиться нова числова характеристика - номер класу, до якого він був віднесений.

На другому етапі проводиться перетворення результатів багатовимірної кластеризації в формат геоданих, підтримуваний в геоінформаційній системі, з метою відображення об'єктів дослідження на цифровій карті з урахуванням їх приналежності до виявлених на першому етапі класів (по номеру класу). Так, об'єкти, що належать певному класу, відображаються, на цифровій карті однаковими умовними позначеннями одного кольору (наприклад, червоні квадрати). Для відображення об'єктів, які відносяться до іншого класу, використовуються інші умовні позначення та / або інший колір (наприклад, зелені трикутники або сині ромби). Відображення об'єктів дослідження в геоінформаційній системі з урахуванням їх приналежності до виділених класів дозволяє виявити закономірності просторового розміщення однорідних об'єктів (об'єктів потрапили в один клас) і визначити межі територіальних зон, в яких вони (однорідні об'єкти) розташовуються.

На третьому етапі проводиться ГІС-аналіз об'єктів дослідження нанесених на цифрову карту. На даному етапі використовуються просторові характеристики об'єктів дослідження і значення номера класу, до якого був віднесений кожен об'єкт на першому етапі. Засобами геоінформаційних систем з використанням методів просторового аналізу здійснюється визначення меж однорідних територіальних зон в межах яких знаходяться об'єкти, що належать до одного класу. При цьому об'єкти, що потрапили в різні класи, будуть розташовуватися в різних територіальних зонах. Сукупність виявлених на даному етапі однорідних територіальних зон і

являють собою імітаційну модель просторової структури розташування об'єктів на досліджуваній території. Імітаційна модель просторової структури розподілених об'єктів в [101] називається геоімітаційною моделлю.

Таким чином, основною особливістю розробленої концептуальної схеми побудови математичної моделі однорідних територіальних зон є поєднання методів багатовимірної кластеризації, які дозволяють розділити всі об'єкти на однорідні класи (групи), і просторового аналізу, який дозволяє виявити межі однорідних територіальних зон на основі результатів кластеризації.

3.8. Алгоритм нечіткої самоорганізації к-середніх

Припустимо, що в мережі існує K нечітких нейронів з центрами в точках c_i ($i = 1, 2, \dots, K$). Початкові значення цих центрів можуть бути обрані випадковим чином з областей допустимих значень відповідних компонентів векторів x_j ($j = 1, 2, \dots, p$), використаних для навчання. Нехай функція фазифікації задана в формі узагальненої функції Гаусса.

Вектор x_j , що подається на вхід мережі, буде належати до різних груп, що представлені c_i , зі ступенем u_{ij} , причому $0 < u_{ij} < 1$, а сумарний ступінь приналежності до всіх груп, очевидно, дорівнює 1. Тому

$$\sum_{i=1}^K u_{ij} = 1 \quad (3.97)$$

для $j = 1, 2, \dots, p$. Функцію похибки, відповідну такому представленню, можна визначити як суму часткових похибок приналежності до центрів c_i з урахуванням ступеня приналежності u_{ij} . Відповідно,

$$E = \sum_{i=1}^K \sum_{j=1}^p u_{ij}^m \|c_i - x_j\|^2, \quad (3.98)$$

де m – це ваговий коефіцієнт, який приймає значення з інтервалу $(1, \infty)$. Мета навчання з самоорганізацією полягає в такому підборі центрів c_i , щоб для заданої множини навчальних векторів x_j – забезпечити досягнення мінімуму функції (3.98)

при одночасному дотриманні умов обмеження (3.97). Таким чином виникає задача мінімізації нелінійної функції (3.98) з p обмеженнями типу (3.97). Вирішення цього завдання можна звести до мінімізації функції Лагранжа, визначеної у вигляді [28].

$$LE = \sum_{i=1}^K \sum_{j=1}^N u_{ij}^m \|c_i - x_j\|^2 + \sum_{j=1}^p \lambda_j \left(\sum_{i=1}^K u_{ij} - 1 \right), \quad (3.99)$$

де λ_j ($j=1, 2, \dots, p$) – це множинники Лагранжа. В [74] доведено, що рішення задачі можна представити у вигляді

$$c_i = \frac{\sum_{j=1}^p u_{ij}^m x_j}{\sum_{j=1}^p u_{ij}^m}, \quad (3.100)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^K \left(\frac{d_{ij}^2}{d_{kj}^2} \right)^{\frac{1}{m-1}}}, \quad (3.101)$$

де d_{ij} – це евклідова відстань між центром c_i , і вектором x_j , $d_{ij} = \|c_i - x_j\|$.

Оскільки точні значення центрів c_i , на початку процесу не відомі, алгоритм навчання має бути ітераційним. Він може бути сформульований в наступному вигляді:

1. Виконати випадкову ініціалізацію коефіцієнтів u_{ij} , вибираючи їх значення з інтервалу $[0, 1]$ таким чином, щоб дотримувалася умова (3.97).
2. Визначити K центрів c_i , відповідно до (3.100).
3. Розрахувати значення функції похибки відповідно до виразу (3.98). Якщо її значення виявиться нижче встановленого порогу або якщо зменшення цієї похибки щодо попередньої ітерації дуже мале, то завершити обчислення. Останні значення центрів складають рішення. В іншому випадку перейти до п. 4.
4. Розрахувати нові значення u_{ij} за формулою (3.101) і перейти до п. 2.

Таку процедуру нечіткої самоорганізації будемо називати алгоритмом C-means.

Багаторазове повторення ітераційної процедури веде до досягнення мінімуму функції E , який необов'язково буде глобальним мінімумом. Якість центрів, що оцінюється значенням функції похибки E , істотно залежить від попереднього підбору як значень u_{ij} , так і центрів c_i . Найкращим може бути визнано таке розміщення центрів, при якому вони розташовуються в областях, що містять найбільшу кількість пред'явлених векторів x_j . При такому підборі центрів вони представлятимуть вектори даних x_j з найменшою сумарною похибкою.

Тому початок ітераційної процедури розрахунку оптимальних значень центрів повинно передувати процедурою їх ініціалізації. До найбільш відомих алгоритмів ініціалізації відносяться алгоритми пікового групування і різницевого групування даних.

3.9. Розробка алгоритму побудови моделі сукупності однорідних територіальних зон

Побудова (створення) імітаційної моделі, відповідно до наведеної в [212, 356] класифікації математичних моделей в залежності від методів реалізації, виконується алгоритмічними методами. Це означає [25], що для реалізації імітаційної моделі створюється спеціальний алгоритм, який визначає структуру і логіку взаємодії елементів у досліджуваній системі. Такий алгоритм в [102] називається-моделюючим. На основі концептуальної схеми в ході дисертаційних досліджень був розроблений алгоритм побудови математичної моделі, однорідних територіальних зон, який можна застосовувати в умовах малого обсягу даних. Алгоритм реалізується з використанням нового підходу, заснованого на поєднанні методів багатовимірної кластеризації і ГІС-аналізу. Розглянемо більш докладно основні етапи алгоритму, що реалізує математичну (імітаційну) модель сукупності однорідних територіальних зон.

Вибір істотних показників. Визначається перелік показників, що характеризують об'єкти дослідження. Проводиться вибір найбільш істотних

непросторових показників, з використанням яких буде побудована імітаційна модель.

Нормування різнорідних багатовимірних даних. Як правило, багатовимірні дані про об'єкти дослідження, отримані в ході натурного експерименту, описуються різнорідними характеристиками.

Тому нормування таких експериментальних даних є необхідним етапом перед їх кластеризацією.

Визначення числа класів. Число класів визначається виходячи з аналізу цілей рішення, завдання та зручності подальшої інтерпретації виявлених територіальних зон. Найбільш часто на практиці виникають завдання виділення зон, в межах яких досліджувані величини мають великі, середні та малі значення. У цьому випадку зручно розділяти сукупність експериментальних даних на 3 класи.

Кластеризація багатовимірних даних. Проводиться поділ сукупності досліджуваних об'єктів, на однорідні класи з використанням алгоритмів багатовимірної кластеризації даних.

Кластеризація проводиться на основі нормованих значень непросторових характеристик об'єктів: дослідження від вибору алгоритму багатовимірної кластеризації буде залежати, результат моделювання та адекватність побудованої імітаційної моделі ОТЗ. Вибір алгоритму кластеризації забезпечує прийнятну якість моделі, розглянуто більш докладно в розділі 4.

Перетворення результатів кластеризації в формат геоданих.

Проводиться перетворення результатів кластеризації в формат, що сприймається геоінформаційними системами. Є важливим сполучним етапом, який забезпечує поєднання методів багатовимірної кластеризації і просторового аналізу.

Створення цифрової карти території засобами ГІС. У разі відсутності готової цифрової карти досліджуваної території, здійснюється її створення засобами, геоінформаційної системи з використанням, різних векторних шарів і даних дистанційного зондування Землі.

Відображення результатів кластеризації на цифровій карті.

Результати кластеризації, перетворені до формату геоданих, з урахуванням просторових властивостей (географічних координат, розмірів та ін.). Пункти відбору проб відображаються на цифровій карті. При відображенні об'єктів, що належать різним класам, на цифровій карті використовуються різні умовні позначення та / або колірні схеми. Побудова моделі сукупності ОТО. З використанням функцій просторового аналізу, реалізованих в геоінформаційній системі, здійснюється виявлення місця розташування і конфігурації однорідних територіальних зон на досліджуваній території. Визначення меж однорідних зон проводиться на основі числового значення номера класу, що позначає приналежність просторових точок до того чи іншого класу.

Дослідження адекватності моделі сукупності ОТО. Проводиться дослідження адекватності реалізованої моделі, що відображає просторове розміщення сукупності однорідних територіальних зон. Модель визнається адекватною, якщо результати моделювання відповідають критеріям дослідника [25,29,30]. У разі, коли побудована модель ОТО визнається неадекватною, то вирішується питання про зміну числа класів, або ще раз проводиться вибір істотних показників, що характеризують об'єкти дослідження. Наведений вище алгоритм побудови моделі сукупності однорідних територіальних зон є універсальним і призначений для вирішення завдань, що відносяться до різних галузей наук про Землю, в яких експериментальні дані представлені багатовимірними вибірками малого обсягу.

3.10. Кластеризація та регіоналізація геопросторових даних в управлінні

3.10.1. Задачі управління екологічними небезпеками

Для того щоб групувати геопросторові об'єкти потрібний спосіб обчислення ступеня близькості між об'єктами, тобто міри відстані (дистанції) між двома довільними об'єктами A і B . Цю міру відстані позначимо як $dist(A,B)$. Для обчислення цієї величини набір атрибутів конкретного об'єкта кластеризації подають часто у вигляді вектора відповідного простору [153].

Властивості міри відстані:

для всіх об'єктів A і B : $dist(A,B) > 0$;

$$\text{dist}(A,B) = \text{dist}(B,A);$$

для кожного з об'єктів A : $\text{dist}(A,A) = 0$;

$$\text{dist}(A,C) < \text{dist}(A,B) + \text{dist}(B,C).$$

Виходячи з обчисленої величини $\text{dist}(A,B)$, можна визначити, наскільки елементи A і B «близькі» один до одного: чим менша відстань, тим «ближчі» елементи, і навпаки [153]..

Часто в кластеризації паралельно з терміном *близькість* використовується термін «схожість» (*similarity*). Міра схожості $\text{sim}(A,B)$ – протилежність мірі відстані.

У кластерному аналізі використовуються різні міри близькості (схожості). В інтелектуальному аналізі найбільш відомі такі міри відстані:

- *відстань Матхатмана*

$$\text{dist}(X,Y) = |x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|, \quad (3.102)$$

де $X = (x_1, x_2, \dots, x_n)$; $Y = (y_1, y_2, \dots, y_n)$; два об'єкти з набором значень атрибутів, які мають бути нормалізовані;

- *евклідова відстань*

$$\text{dist}(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}, \quad (3.103)$$

- *схожість за косинусом*

$$\text{dist}(X,Y) = 1 - \text{sim}(X,Y), \quad (3.104)$$

$$\text{де } \text{sim}(X,Y) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}}. \quad (3.105)$$

Ця формула визначає косинус кута між векторами X і Y . В інформаційно-пошукових системах ця величина визначає ступінь близькості (коефіцієнт кореляції) пошукових образів X і Y [153].

Метод неієрархічної кластеризації K-means являє собою спробу визначити оптимальну кількість місць K , де сума відстаней від кожної точки до кожного з центрів K зведено до мінімуму, що називається глобальною оптимізацією. На

практиці реалізуються початкові припущення про місця K і здійснюється локальна оптимізації для місць кластерів щодо сусідніх точок. Таким чином, дві K -means процедури можуть не виробляти ті ж результати, навіть якщо K є ідентичними із за кількох основних методів локальної оптимізації [153].

Алгоритм K -середніх побудований на чотирьох основних операцій:

- вибір початкових точок для K кластерів;
- розрахунок несхожості між об'єктом і середнім значенням кластера;
- включення об'єкта в кластер, у якого середнє ближче до об'єкта;
- перерахунок середнього значення кластера із об'єктів, які належать

йому, щоб всередині кластера несхожість зводилась до мінімуму. Крім першої операції інші три операції виконується повторно в алгоритмі до виконання умови (поки немає більше точок, що змінюють кластери). Суть алгоритму полягає в мінімізації функції вартості, яка залежить від несхожості між мірою кожного спостереження з середнім значенням кластеру. Несхожість, як правило, моделюється як евклідова відстань в k -means. Функція витрат полягає в наступному:

$$\text{Minimize} \sum_{j=1}^n \sum_{k=1}^k a_j d_{jk} z_{jk}, \quad (3.106)$$

де j, k позначає загальну кількість спостережень і кластерів;

a_j позначає вагу спостереження j ;

d_{jk} позначає відстань між спостереженням j і центром кластера k ;

$$z_{jk} = \begin{cases} 1, & \text{якщо спостереження } j \text{ є в кластері } k, \\ 0, & \text{в іншому випадку} \end{cases}, \quad (3.107)$$

z_{jk} є індикатором приналежності спостереження до кластеру, за умови, що z_{jk} приймає значення 1 тільки один раз для призначеного кластера, і значення 0 для інших кластерів [153].

Неієрархічні методи кластеризації краще підходять до аналізу геопросторових даних, тому що дозволяють об'єднувати об'єкти одразу у географічні групи. В той

же час ієрархічні методи мають більш розвинутий механізм для визначення кількості кластерів, що застосовується у дослідженні.

Для геоінформаційних продуктів притаманний одно- або двовимірний аналіз векторних або растрових моделей даних, коли дані групуються на основі одного чи двох параметрів та відображаються за допомогою картограми. З іншого боку, статистичні пакети мають досконалі модулі кластерного аналізу, але при цьому відсутня можливість автоматизованого картографування даних та використання географічної інформації для цілей поділу об'єктів на групи. В останній час ця проблема вирішується за рахунок інтеграції статистичних пакетів та геоінформаційних продуктів через протоколи взаємодії даних (експорт / імпорт даних, підтримка форматів), але така процедура вимагає наявності декількох програмних продуктів на робочому місці та досвідченого користувача. Тому існує потреба у створенні додаткових модулів кластерного аналізу, які працюють безпосередньо у геоінформаційному середовищі та використовують його переваги.

Такий програмний модуль було розроблено у вигляді інструменту кластеризації на основі розширень геоінформаційної системи ArcGis 10x [153].

Алгоритм використання програмного модулю зображено на рис. 3.1. До основних його переваг відноситься можливість проведення операцій кластеризації у багатовимірному просторі, залучення географічної інформації як додаткових координатних вимірів об'єкту, побудова автоматизованих картографічних моделей та легенд за результатами аналізу. Використання географічних координат дозволяє врахувати взаємне розташування об'єктів у географічному просторі, топологію їх зв'язків та відповідно здійснювати районування території на основі математично обґрунтованих даних [128].

Програмний модуль розроблено за допомогою мови C#, набору засобів розробки ArcGis SDK для .NET архітектури. Модуль містить набір методів агрегації даних на основі методу ближнього сусіда, Уорда, середньої відстані, центральної відстані, дальнього сусіда, а також можливість задавати кількість вихідних кластерів. Результатом роботи модуля є тематична карта та легенда до неї

з виділеними кластерами, таблиця з розбивкою за класами географічних об'єктів, дендрограма кластерного аналізу (рис. 3.1) [153].

Для проведення кластеризації було обрано адміністративно-територіальний поділ України на 24 області, Автономну Республіку Крим та міста республіканського підпорядкування Київ та Севастополь. В якості вхідних даних для кластеризації використано статистичні дані за 2013 р. про утворення та накопичення на території регіонів відходів 1-4 класів небезпеки, що дало можливість проведення аналізу у 8-вимірному просторі [17]. Основним завданням кластеризації за цими показниками є виявлення основних типів регіонів за особливостями утворення відходів та небезпек з ними пов'язаних.

Основними характерними особливостями розподілу відходів в Україні є їх накопичення у регіонах видобутку корисних копалин та промислового виробництва у східних, центральних та деяких західних областях України. Пов'язані з цим небезпеки залежать від обсягів утворення та накопичення відходів. Відходи 1 групи небезпеки накопичені переважно у Донбасі, а також Кіровоградській, Сумській та Харківській областях, а за їх утворенням перші місця займають Луганська, Харківська, Полтавська та Донецька області. Відходи 2 групи накопичені у АР Крим, Донецькій та Луганській області, а за їх утворенням очолюють список АР Крим, Дніпропетровська, Сумська та Донецька області. Найбільша кількість відходів 3 групи накопичена у Запорізькій та Донецькій області, а за їх щорічним утворенням перші місця займає Полтавська, Миколаївська та Донецька області. Абсолютні обсяги накопичення відходів 4 класу небезпеки зафіксовані у Дніпропетровській, Донецькій та Луганській областях, які разом з Кіровоградщиною також мають найбільші обсяги щорічного утворення відходів цієї групи [128].

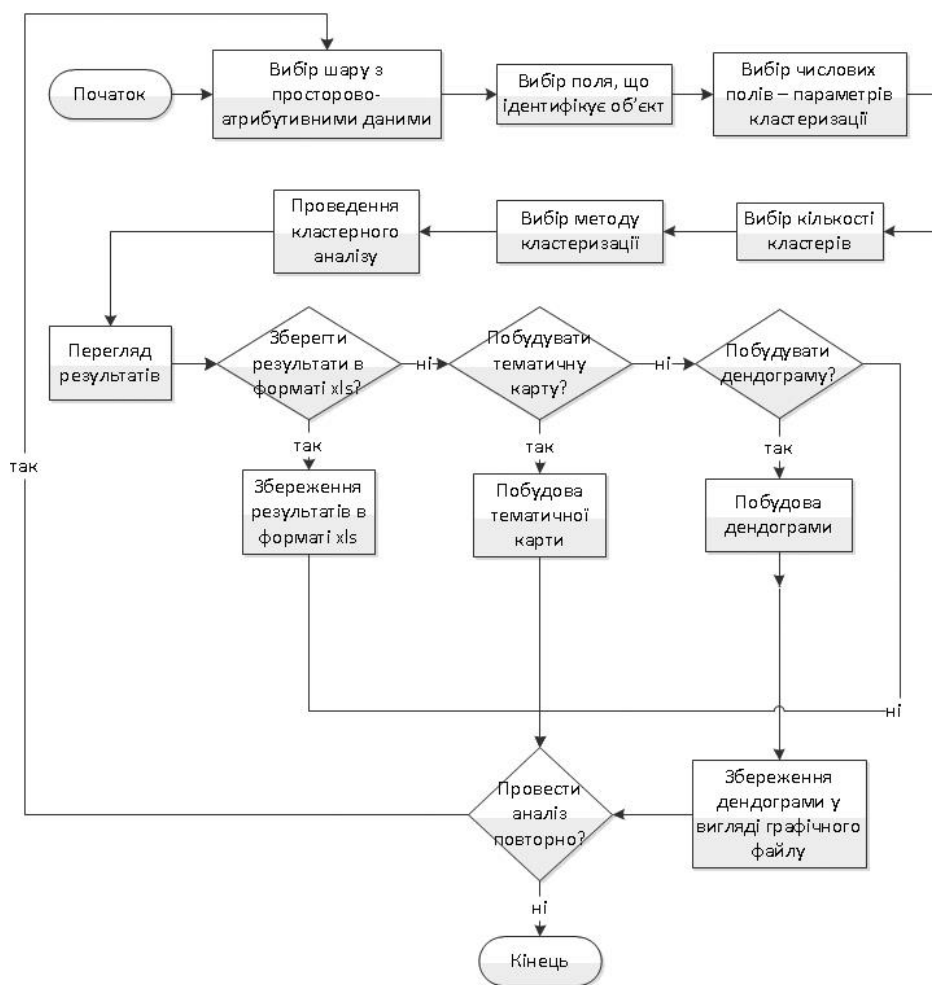


Рис. 3.1. Алгоритм використання модуля кластерного аналізу

Таблиця 3.3

Кластеризація регіонів із врахуванням геопросторових даних
(метод – відстань Уорда)

Кластер	1	2	3	4	5	6	7
Об'єкти	Волинська	Київська	Донецька	Чернігівська	Полтавська	Дніпропетровська	Автономна Республіка Крим
	Рівненська	м. Київ	Луганська	Тернопільська	Харківська	Запорізька	Львівська
	Житомирська	Черкаська		Хмельницька	Миколаївська		Івано-Франківська
	Закарпатська	Кіровоградська		Сумська	Херсонська		

	Чернівецька			Одеська	М. Севастопол ь		
				Вінницька			

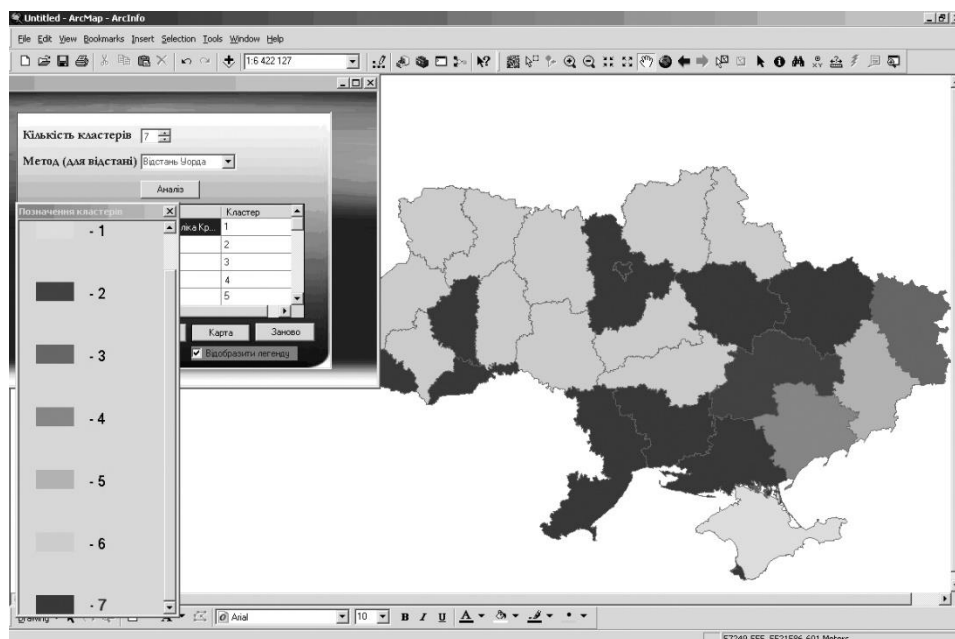


Рис. 3.2. Вікно ArcGis з результатами кластерного аналізу

Після порівняльного аналізу методів кластеризації було вирішено обрати метод Уорда, який ґрунтується на принципі мінімізації дисперсії двох кластерів, що об'єднуються на кожному кроці [169].

$$V_l = \sum_i \sum_j (x_{ij} - x_{jl})^2, \quad (3.108)$$

де l - номер кластера;

i – номер об'єкта ($i=1,2, \dots, n_l$);

n_l – кількість об'єктів в l -том кластері;

J – номер ознаки ($j = 1,2,\dots,k$);

k – кількість ознак, що характеризують кожний об'єкт.

В дослідженні аналізовано випадок кластеризації у 8- та 10-вимірному просторі. В другому випадку використано в якості додаткових показників географічні

координати центрів регіонів, які дозволяють врахувати близькість розміщення адміністративних одиниць та провести їх зонування [128].

У 8-вимірному просторі виділяється 5 кластерів регіонів. Перший кластер містить АР Крим, 2 – Дніпропетровська та Луганська область, 3 – Запорізька та Донецька область, 4 - Сумська, Кіровоградська, Івано-Франківська, Чернігівська, Рівненська, Черкаська, Житомирська, Львівська, Хмельницька, Волинська, Вінницька області. 5 – Харківська, Херсонська, Київська, Одеська, м. Київ, Тернопільська, Закарпатська, Чернівецька, м. Севастополь, Полтавська, Миколаївська.

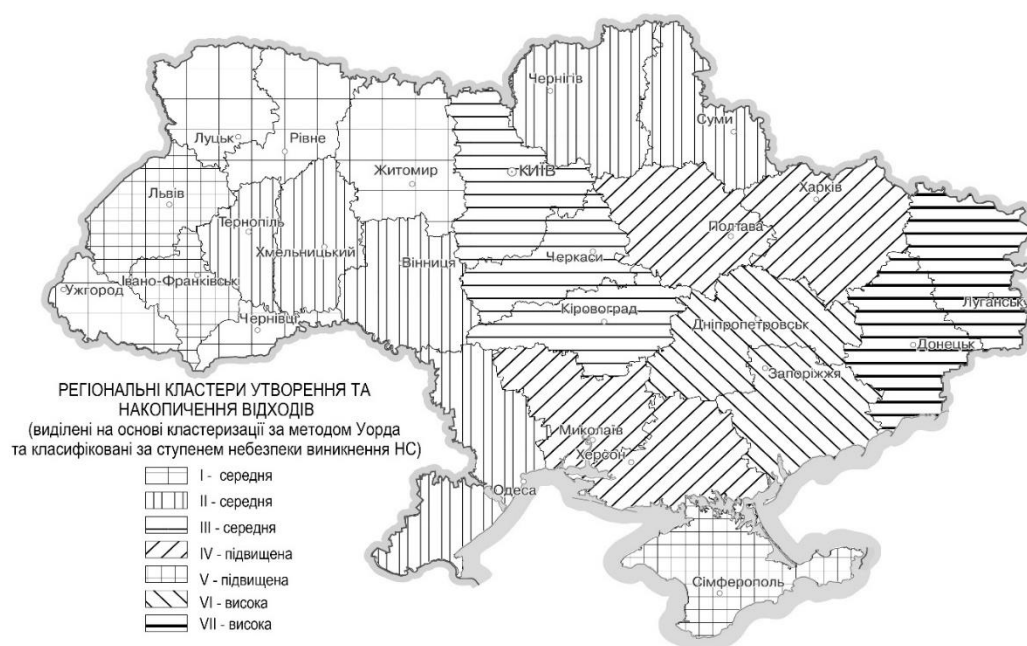


Рис. 3.3. Результати кластеризації за просторово-атрибутивними ознаками поводження з відходами

У 10-вимірному просторі виділяється 7 кластерів, які демонструють географічне та тематичне поєднання (див. табл. 3.3). За цими даними було розроблено карту території України, на якій виділено сім груп областей (рис.3.3). Перший кластер містить Донецьку та Луганську області з високим рівнем накопичення відходів, зокрема відходів 1 та 4 класів, що свідчить про високий ступінь небезпеки пов'язаної з високотоксичними речовинами у місцях їх складування та небезпечними процесами, що пов'язані з відходами 4 класу. До 2

кластеру відноситься Дніпропетровська та Запорізька області з високим ступенем небезпеки хімічних відходів 1 класу та відходів металургії. 3 кластер містить АР Крим, Львівську та Івано-Франківську області, які вирізняються наявністю відходів 2 – 3 класів небезпеки, що пов'язано з видобутком та промисловим виробництвом хімічних сполук. Склад 4 кластеру містить Київ, Київську, Черкаську та Кіровоградську області, де переважають токсичні відходи гірничого видобутку та відходи 1-2 класів небезпеки при хімічному виробництві. 5 кластер містить Харківську, Полтавську, Миколаївську, Херсонську області та м. Севастополь, де переважають відходи 3-4 ступенів небезпеки. 6 кластер складається з Чернігівської, Тернопільської, Хмельницької, Сумської, Одеської, Вінницької областей з відносно низьким рівнем небезпеки відходів, які переважно відносяться до 3 та 4 групи небезпеки. Останній, 7 кластер складається з Волинської, Рівненської, Житомирської на півночі та Закарпатської, Чернівецької областей на заході України з низькими обсягами утворення та складування відходів. Отримані кластери у своїй більшості мають компактну географічну структуру, яка дозволяє проводити не лише типізацію території, але здійснювати районування за певним набором ознак [128].

3.10.2. Регіоналізація території України за результатами кластеризації у просторі показників сталого розвитку

Для проведення кластеризації було обрано кластеризацію регіонів України в 3 та 5 мірному просторі.

Тривимірний простір характеризує якість життя населення як ключову характеристику сталого розвитку, що включає індекс економічного, екологічного та соціально-інституціонального розвитку. П'ятивимірний простір включає категорію політики «Безпека» та гармонізацію складових якості життя.

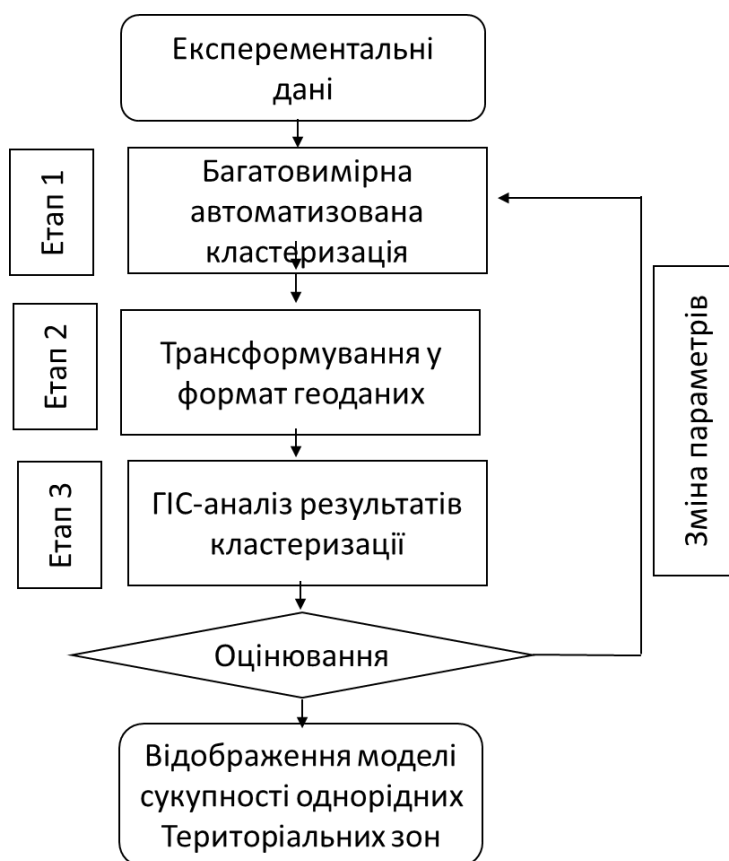


Рис. 3.4. Концептуальна схема побудови математичної моделі сукупності однорідних територіальних зон

Для кожної групи показників були проведені оцінки за основними параметрами просторової кластеризації. В результаті було з'ясовано, що найбільш компактні групи кластеризації можна отримати при використанні методу спільних ребер. Кластеризація за основними параметрами якості життя утворила найбільш компактні кластери при розподілі на шість груп. Кожний кластер описується розподілом на статистичній діаграмі (рис. 3.5 – 3.6).

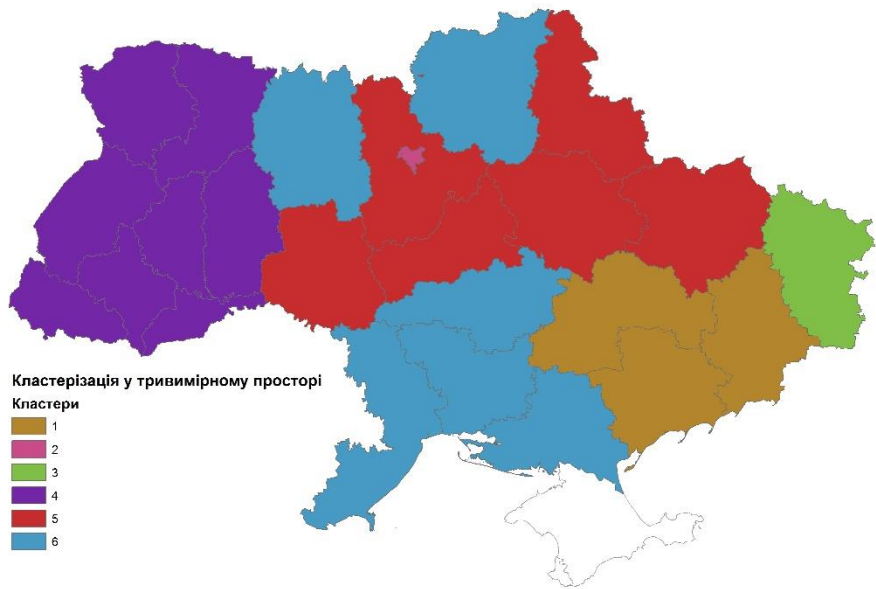


Рис.3.5. Картограма кластеризації регіонів України за індикаторами якості життя (2014 р.)

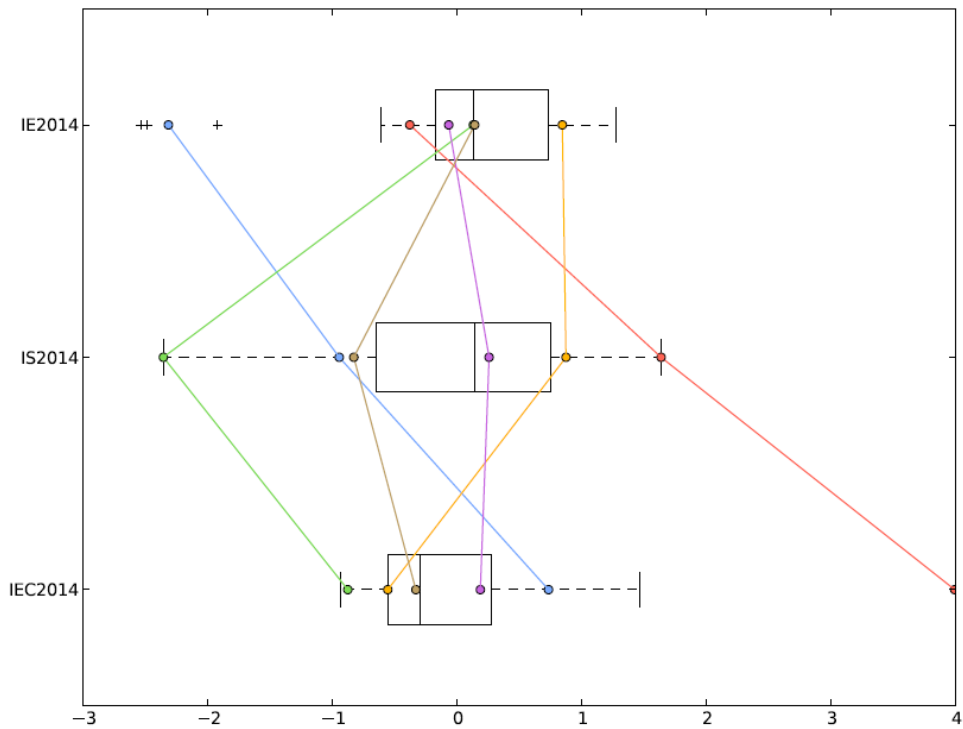


Рис. 3.6. Конфігурація кластерів за якістю життя

Таблиця 3.4. Загальна статистика змінних

Змінна	Середнє	Ст. відх.	Min	Max	R2
--------	---------	-----------	-----	-----	----

IE2014	0,5106	0,0773	0,3143	0,6090	0,8835
IEC2014	0,2620	0,1433	0,1285	0,8339	0,8659
IS2014	0,5494	0,0420	0,4505	0,6181	0,8609

Таблиця 3.5. Показники групування кластерів

Номер кластеру	Стандартна відстань	Сума квадратів відхилень
1	0,0844	2,8445
2	0,0609	2,1538
3	0,0684	2,3340
4	0,0505	2,4736
5	0,0709	2,1038
6	0,0584	2,3220

Перший кластер включає три регіони сходу країни і характеризується низьким рівнем індексу екологічного виміру і порівняно високими іншими складовими. На цьому фоні Луганська область певним чином відрізняється, оскільки статистика збирається тільки з північної частини області, яка має сільсько-господарську орієнтацію. Регіони західної України відносяться до наступного кластеру, який має досить збалансовані вище середнього показники вимірів окрім економічного. Центральна Україна має середні збалансовані показники за всіма параметрами. Південні області, Житомирська та Чернігівська області мають знижені соціально-інституціональні показники.

За всіма видами кластеризації місто Київ потрапляє в окремий кластер, оскільки має найбільшу варіацію значень параметрів.

Кластеризація в п'ятимірному просторі дозволила агрегувати регіони ще більше, утворивши три основні макрорегіони. Найбільший вклад в це здійснив параметр безпеки життєдіяльності (рис. 3.7 – 3.8).



Рис. 3.7. Картограма кластеризації регіонів України за категоріями політики (2014) р.

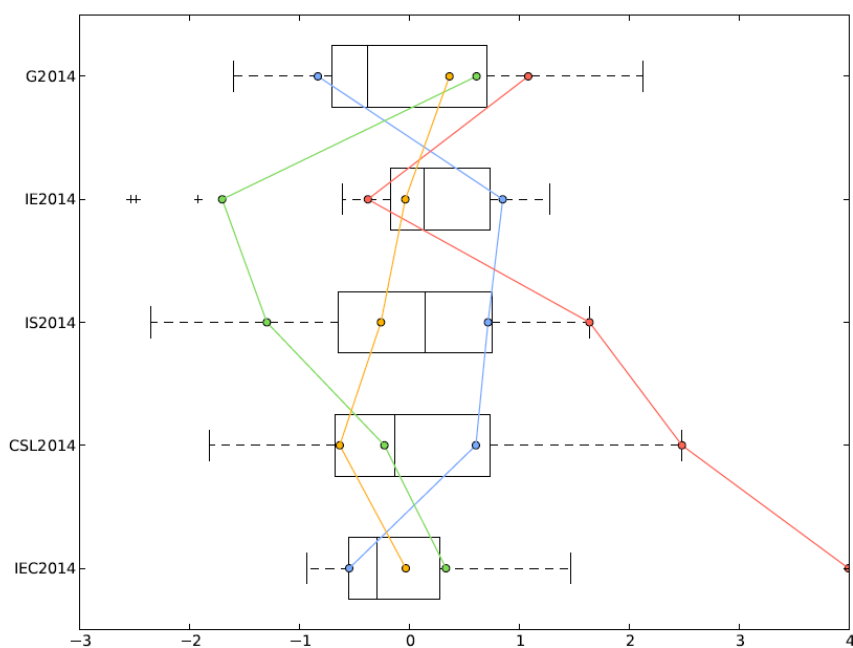


Рис. 3.8. Конфігурація кластерів за категоріями політики

Таблиця 3.6. Загальна статистика змінних

Змінна	Середнє	Ст. відх.	Min	Max	R2
IE2014	0,5106	0,0773	0,3143	0,6090	0,8835
IEC2014	0,2620	0,1433	0,1285	0,8339	0,8659
IS2014	0,5494	0,0420	0,4505	0,6181	0,8609

CSL2014	1,2282	0,1358	0,9807	1,5644	0,5620
G2014	0,6651	0,0955	0,5122	0,8674	0,4142

Таблиця 3.7. Показники групування кластерів

Номер кластеру	Стандартна відстань	Сума квадратів відхилень
1	0.1245	11.7049
2	0,0509	2,1738
3	0.1883	15.8084
4	0.1458,	21.0225

Перший кластер охоплює східні регіони України, має низькі показники екологічного і соціально-інституціонального вимірів. Другий кластер включає центральні, південні та північно-східні регіони України. Він характеризується низькою безпекою та середніми показниками інших параметрів. Третій кластер згруповано на заході країни, який має підвищені параметри екологічного та соціально-інституціонального вимірів та безпеки, але низькі значення економічного виміру та гармонізації. Київ потрапляє в окремий кластер з дуже великим розбігом показників, де лідирує економічний вимір.

3.11. Геопросторовий аналіз даних з використанням методу гарячих точок

3.11.1. Кластеризація геопросторових даних про військову конфліктогенність у світі

В проектах, спрямованих на вивчення, передбачення та попередження виникнення збройних конфліктів, використання засобів розвідувального аналізу є невід'ємною частиною робочого процесу. За допомогою баз даних можна простежити зв'язки та процеси у військових конфліктах, знайти нові фактори, що їх провокують, а також способи для їх запобігання. Використання засобів геовізуалізації даних в якості інструменту для моделювання військових конфліктів, вивчення даних і створення гіпотез, створення інтерактивних карт є сучасним і

ефективним підходом, що спрощує роботу з даними та доступ до них усім зацікавленим сторонам проєктів.

В якості вхідних даних було взято масив від проєкту Uppsala Conflict Data Program Georeferenced Event Dataset (UCDP GED) — набір геокодованих даних про конфлікти від кафедри досліджень питань миру і конфліктів університету Упсала (Швеція) [379,380]. Метою цього проєкту є надання академічній спільноті найбільш повних структурованих даних про події організованого насильства у світі в період з 1989 року, задовольняючи потребу в географічно- і часово-деталізованих даних.

Для кластеризації геопросторових даних було вирішено завдання перетворення даних у формат бази геоданих, геокодування, кластеризація та візуалізація за допомогою сучасних програмних бібліотек. В якості СУБД для зберігання даних було обрано MySQL, середовища обробки геоданих – QGIS.

Кластеризацію даних за геопросторовими ознаками було здійснено на основі методу «гарячі та холодних» точок, що базується на аналізі геопросторової статистики.

Більшість статистичних тестів починаються з визначення нульової гіпотези. Нульова гіпотеза для інструментів аналізу структурних закономірностей — це повна просторова хаотичність або самих об'єктів або значень, пов'язаних з ними. Р-значення — це ймовірність, Z-оцінки є стандартними відхиленнями. Z-оцінки і р-значення, отримані в результаті аналізу структурних закономірностей, свідчать про те, чи можна відхилити нульову гіпотезу чи ні. Як правило, використовують один з інструментів аналізу структурних закономірностей, припускаючи, що z-оцінка і р-значення будуть свідчити про можливе спростування нульової гіпотези. Це буде говорити про те, що досліджувані об'єкти або значення, пов'язані з ними, виявляють статистично значущу кластеризацію або дисперсію. Всякий раз, коли спостерігається просторова структура, така як кластеризація ландшафту (або просторових даних), можна побачити докази роботи деяких основних просторових процесів [156].

Ключова ідея полягає в тому, що значення в середині нормального розподілу представляють очікуваний результат. Коли абсолютне значення z -оцінки є великим, а ймовірності є маленькими (в хвостах нормального розподілу), спостерігається незвичний розподіл. Для методу “Аналіз гарячих точок” це означає статистично істотну “гарячу” або “холодну” точку.

Метод “Гарячі точки” розраховує статистичний показник для кожної події в наборі даних. Підсумкові p -значення (ймовірності) та z -оцінки (стандартні відхилення) говорять про те, в якій області простору кластеризуються події з високими або низькими значеннями [6]. Метод працює шляхом аналізу кожної події в контексті сусідніх за географією подій. Щоб бути статистично суттєвою гарячою точкою, подія повинна мати високе значення і бути оточена іншими подіями з також високими значеннями. Локальна сума для події та її сусідів порівнюється пропорційно із сумою всіх подій; коли локальна сума дуже відрізняється від очікуваної локальної суми, і коли ця відмінність є занадто великою, щоб бути результатом випадкового процесу, отримується статистично значуща z -оцінка.

Метод гарячих точок використовує формули:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{j,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}}, \quad (3.109)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}, \quad (3.110)$$

де x_j — атрибутивне значення для подій j , $w_{i,j}$ — просторова вага між подіями i та j , n — загальна кількість подій, \bar{X} — середнє арифметичне значень подій, S — дисперсія.

Статистична величина G_i^* , для кожної події в наборі даних, є z -оцінкою. Для статистично значущих позитивних z -оцінок, чим більша z -оцінка, тим інтенсивніша кластеризація високих значень (гаряча точка). Для статистично значущих негативних z -оцінок, чим менше z -оцінка, тим інтенсивніша кластеризація низьких значень (холодна точка). На виході отримуємо новий набір

даних про події з z-оцінкою, p-значенням і рівнем достовірності G_i_Bin для кожної події у вхідному масиві.

Для проведення аналізу даних та їх кластеризації за методом “Гарячих точок” було використано набір UCDP GED у форматі Shapefile “ged50-shp” та у вигляді дампу бази даних MySQL “ged50-mysql”. В ході роботи з набором даних було застосовано програмне забезпечення QGIS з відкритим кодом та комерційну програму ArcGIS від компанії ESRI. З інструментарію ArcGIS для отримання кластеризованих даних було використано вбудований інтерпретатор Python, бібліотеку “arcpy”, функцію генерації просторово-часових вагових коефіцієнтів `GenerateSpatialWeightsMatrix_stats` та власне функцію кластеризації за методом гарячих точок `HotSpots_stats`.

Весь набір даних UCDP GED було розбито за ознакою приналежності до одного з 5 регіонів: Європа, Азія, Північна та Південна Америка, Африка, Близький Схід. Для кожного регіону було обраховано матрицю часово-просторових вагів з параметрами часового діапазону у 5 років та територіального віддалення від точок-сусідів до 500 км.

Використовуючи обраховані матриці вагів, було створено нові набори даних для відповідних регіонів, що містять інформацію про приналежність до гарячих або холодних точок.

Отримані дані було експортовано в окремі шейп-файли, після чого за допомогою вільного програмного забезпечення QGIS імпортувались до бази MySQL. Результуюча база даних має 2 таблиці: власне таблиця даних UCDP GED “ged50-mysql” та таблиця обрахованих z-оцінок, p-значень та значення рівнів достовірності розподілу G_i_Bin . Ці таблиці зв'язані за полем унікального ідентифікатора події `id` (рис. 3.9, 3.10).

Для візуалізації результатів аналізу та надання доступу до них через мережу Інтернет, на клієнтській частині використовуються засоби відкритого проекту WebGL Globe та бібліОТОки візуалізації Three.js. Це дозволяє завантажувати дані з серверу у форматі JSON та забезпечувати інтерактивну взаємодію користувача з ними.

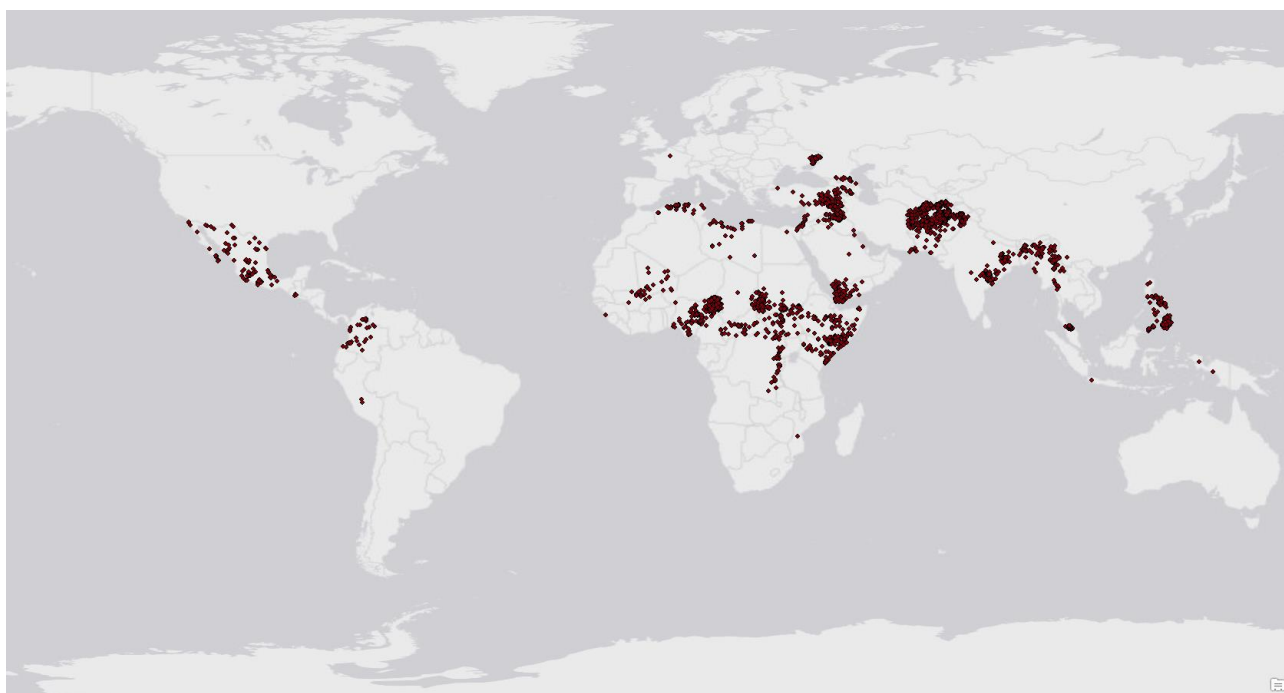


Рисунок 3.9. Розподіл військових конфліктів у світі, 2015 р.

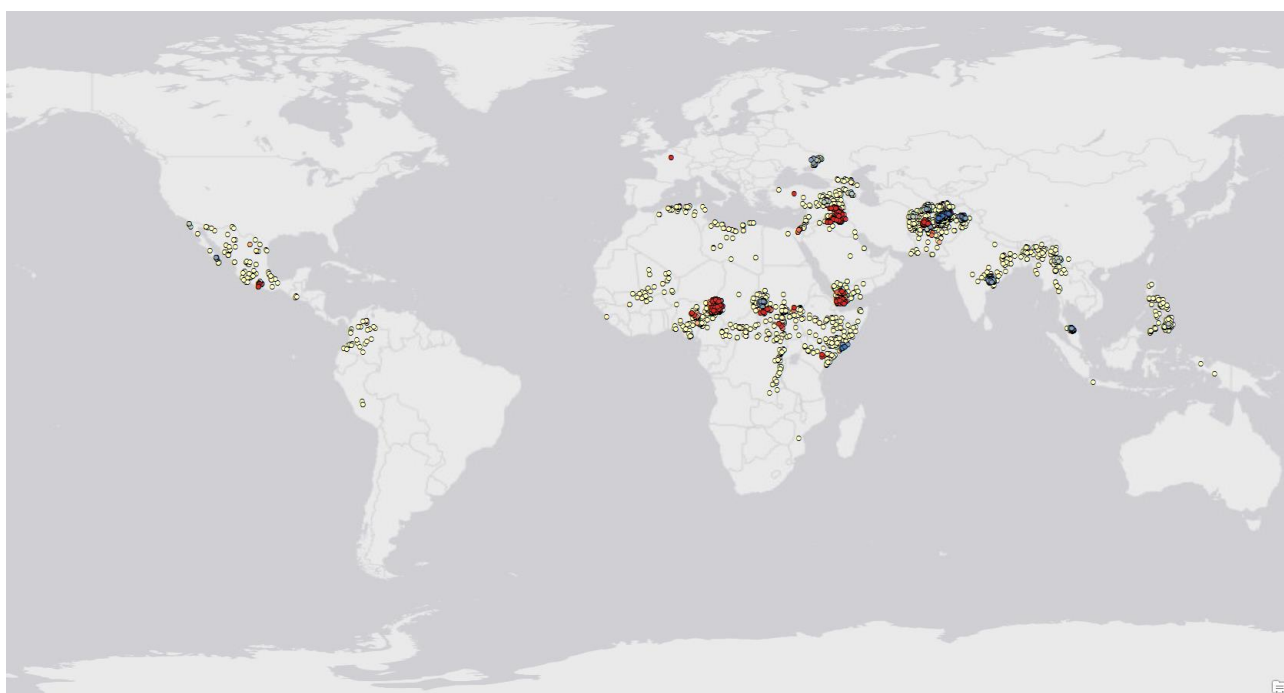


Рисунок 3.10. Кластеризовані військових конфліктів у світі, 2015 р.

На стороні серверу використано СУБД MySQL для збереження результуючого набору даних, веб-сервер Apache для обробки асинхронних запитів клієнтської частини та власне програмне забезпечення на мові PHP, що забезпечує бізнес-логіку роботи веб-додатку. Вигляд додатку приведено на рисунку 3.11.

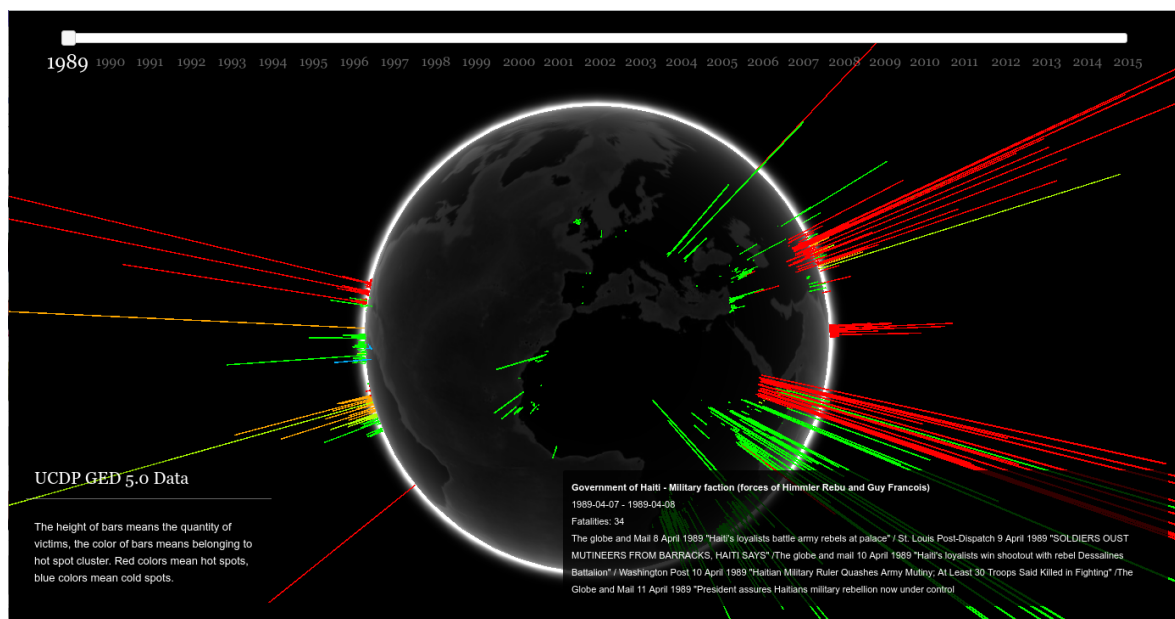


Рисунок 3.11. Вигляд веб-додатку візуалізації результатів аналізу на 3D-глобусі

Для полегшення взаємодії користувача з додатком спосіб перемикання між роками було реалізовано у вигляді повзунка. Самі події представлені у вигляді геоприв'язаних кольорових стовпців різної величини. Висота стовпця відображає кількість жертв конфлікту, а колір — приналежність до гарячих або холодних точок. При наведенні курсора на стовпець в правому нижньому кутку екрану з'являється коротка довідка про дану подію.

Інтерпретація результатів кластеризації може допомогти у виявленні просторово-часових закономірностей виникнення та розвитку конфліктів. Застосований підхід дозволяє визначити «гарячі точки» конфліктів, що можуть слугувати вказівками на особливу небезпеку розвитку інцидентів в регіоні. «Холодні точки», навпаки, можуть вказувати на можливості швидкого згортання ескалації. Аналіз даних в динаміці за останні 30 років дозволяє більш детально розглянути перебіг військових конфліктів у регіонах світу.

3.11.2. Аналіз динаміки вегетаційного стану рослинності

Сервіси земного покриття Copernicus дозволяють провести багатостороннє дослідження змін сезонної вегетації із використанням низки вегетаційних індексів.

До найбільш значущих індексів можна віднести індекси NDVI, LAI, VPI, VCI.

NDVI - нормалізований відносний індекс рослинності - простий кількісний показник кількості фотосинтетично активної біомаси (зазвичай має назву вегетаційний індекс). Один з найпоширеніших і використовуваних індексів для вирішення завдань, що застосовують кількісні оцінки рослинного покриття [162].

Обчислюється за такою формулою:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (3.111)$$

де,

NIR - відображення в ближній інфрачервоній області спектра

RED - відображення в червоній області спектра

Відповідно до цієї формули, щільність рослинності (NDVI) в певній точці зображення дорівнює різниці інтенсивностей відбитого світла в червоному і інфрачервоному діапазоні, поділений на суму їх інтенсивностей. Відповідно до цього щільність рослинності (NDVI) в певній точці зображення дорівнює різниці інтенсивності відбитого світла в червоному і інфрачервоному діапазоні, поділений на суму їх інтенсивності [162].

Низькі значення NDVI вказують на поганий стан рослинного покриття, що може бути викликано посухою і відповідно призвести до підвищеної пожежної небезпеки. В даному випадку важливо зберігати відповідний масштаб та роздільну здатність даних, які повинні бути більш узагальненими ніж у випадку звичайного сільськогосподарського моніторингу [162].

LAI (Leaf Area Index) – Індекс листової площі, який визначається як половина загальної площі зелених елементів купола на одиницю горизонтальної ділянки. Значення отримане із супутника відповідає загальному зеленому LAI всіх шарів пологу лісу, в тому числі підліску, які можуть становити дуже важливий внесок для лісів. Практично LAI вимірює товщину рослинного покриття. LAI визнається в якості однієї з найважливіших кліматичних змінних (ECV) в рамках Глобальної системи спостережень за кліматом (GCOS).

VPI (Vegetation Productivity Index) – Індекс продуктивності рослинності оцінює загальний стан рослинності шляхом посилення на поточне значення NDVI у порівнянні з довгостроковою статистикою за той же період. VPI є процентиллю ранжування поточного значення NDVI у порівнянні зі своїм історичним діапазоном мінливості: значення 0%, 50% і 100% відповідно, свідчать про те, що поточне спостереження співвідноситься з історичним мінімумом (найгірший стан рослинності), медіаною (нормальний) або максимумом (найкраща ситуація), що коли-небудь спостерігалися.

VCI (Vegetation Condition Index) – Індекс стану рослинності порівнює поточний NDVI в діапазоні значень, які спостерігалися в той же період в попередні роки. VCI виражається у відсотках і дає уявлення про те, де спостережуване значення знаходиться між крайніми значеннями (мінімум і максимум) в попередні роки. Нижчі і вищі значення вказують, відповідно, на погані і добрі умови вегетаційного стану.

Основними методологічними засадами розробки динамічних моделей є визначення усереднених показників за певний час спостережень, відносно яких відбувається аналіз. Чим більшим буде ряд спостережень, тим більш якісними можуть бути отримані результати. З іншої сторони довший ряд спостережень дозволяє спрогнозувати глобальні та регіональні тренди з високим ступенем вірогідності. Сукупність індексів оцінки вегетації не дає змоги вибору оптимального показника, який мав би найбільшу кореляцію з регіональними змінами клімату. Тому вибір оптимальних показників залишається складною науковою задачею. Проте, підхід до сумісного аналізу даних різної природи може допомогти у вирішенні цього завдання. Для цього необхідно вирішити базові завдання зонування, нормалізації, зважування, та сумісної обробки отриманих рядів статистичних даних.

Для розрахунку інтегрального індексу за показниками вегетаційних індексів пропонується формула, яка дозволяє поєднати усі різниці міжсезонної зміни індексів для визначання регіональних трендів:

$$I = \frac{\sum_N \sum_L (NDVI_t - NDVI_{t-1}) / N \sum_L (LAI_t - LAI_{t-1}) / L \sum_C (VCI_t - VCI_{t-1}) \sum_P (VPI_t - VPI_{t-1})}{n}, \quad (3.112)$$

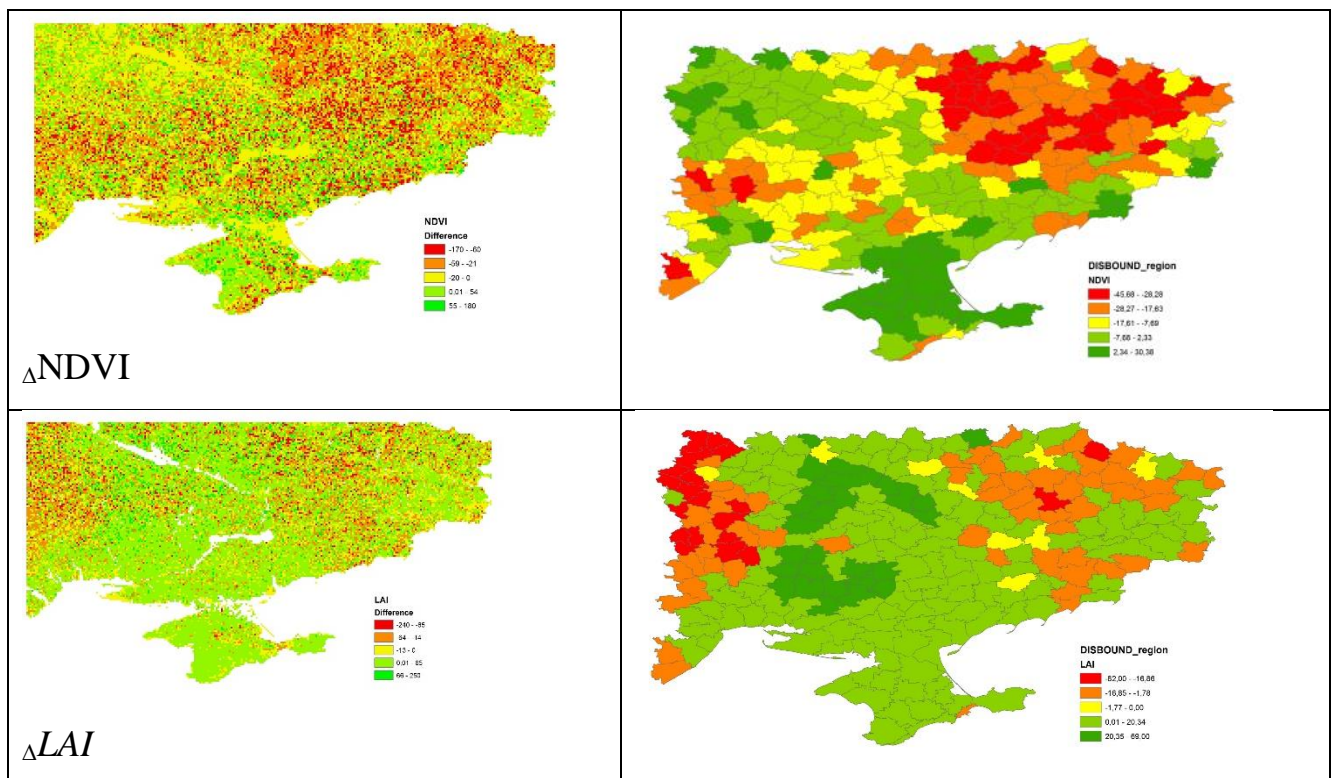
де I - інтегральний індекс,

N – кількість продуктів, які приймають участь у розрахунку,

N, L, C, P – кількість результатів класифікації знімків за індексами [162].

Для проведення експериментальних розрахунків були використані дані вегетаційних індексів за період червень 2014 – червень 2015 років для південно-східної території України, яка одночасно є регіоном високо розвинутого землеробства з високим експортним ПОТОНціалом, а також значною частиною знаходиться у зоні ризикованого землеробства.

Загальний процес геоінформаційної обробки даних включає завантаження відповідних індексів з порталу, маскуванню території з використанням кордону України, обрахунок попиксильної різниці між даними за різний період. На наступному етапі отримуємо зональну статистику для кожної адміністративної одиниці окремо. Дані в такому вигляді простіше нормалізуються та можуть бути використані безпосередньо у процесі підтримки прийняття рішень місцевими органами влади (рис. 3.12).



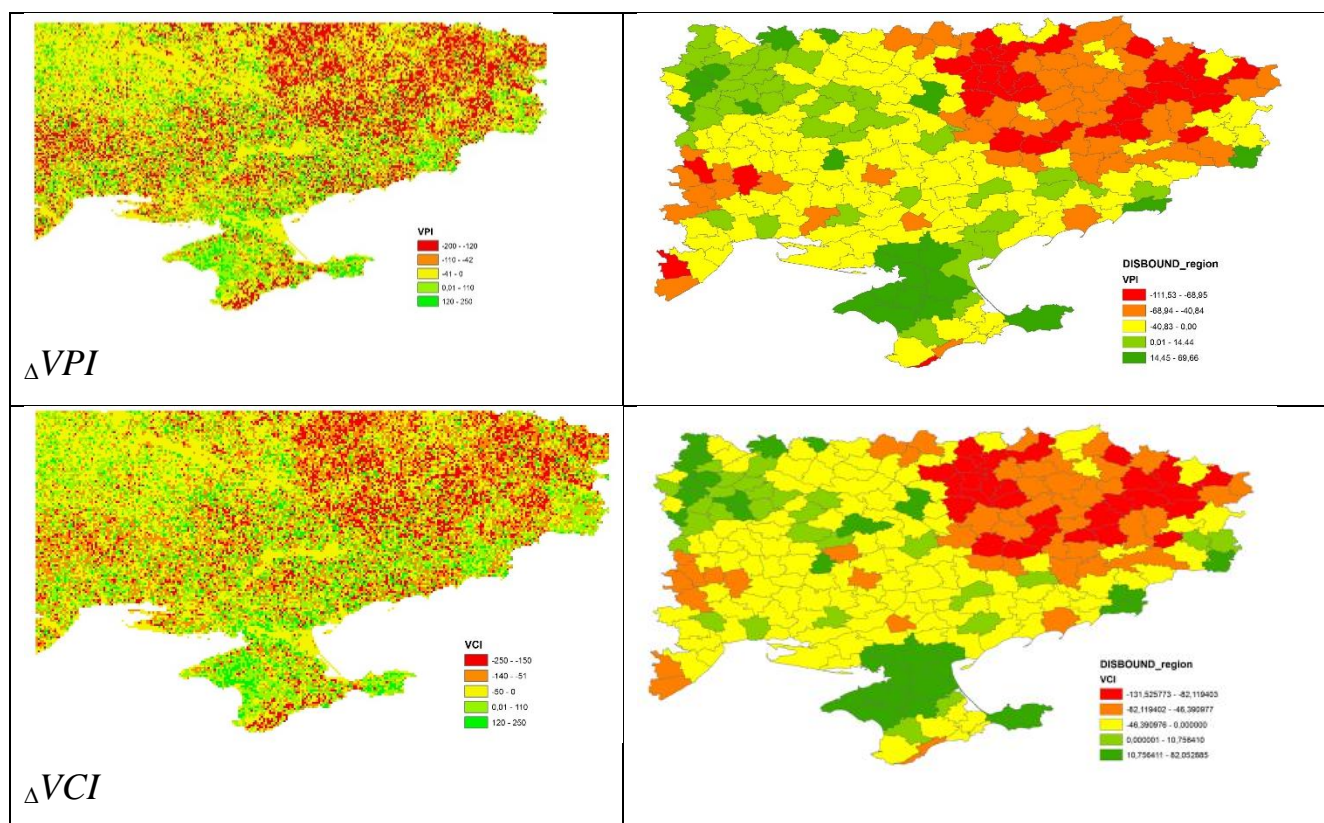


Рис.3.12. Розподіл значень різниці вегетаційних індексів та усереднена зональна статистика за адміністративними районами [162].

Результати інтегрального індексу демонструють наявність виражених регіональних змін, які прослідковуються навіть для річної динаміки даних. Чітко видно, що показники вегетаційних індексів для сходу країни погіршилися, тоді як для півдня та центральної частини України вони дещо покращилися.

З метою перевірки гіпотези регіоналізації було використано метод гарячих точок, який дозволяє визначити просторові кластери стійких позитивних або негативних змін.

Використання методу «Гарячих точок» дозволило отримати статистично значущі просторові кластери високих та низьких значень у розподілі інтегрального індексу вегетації (рис.3.13).

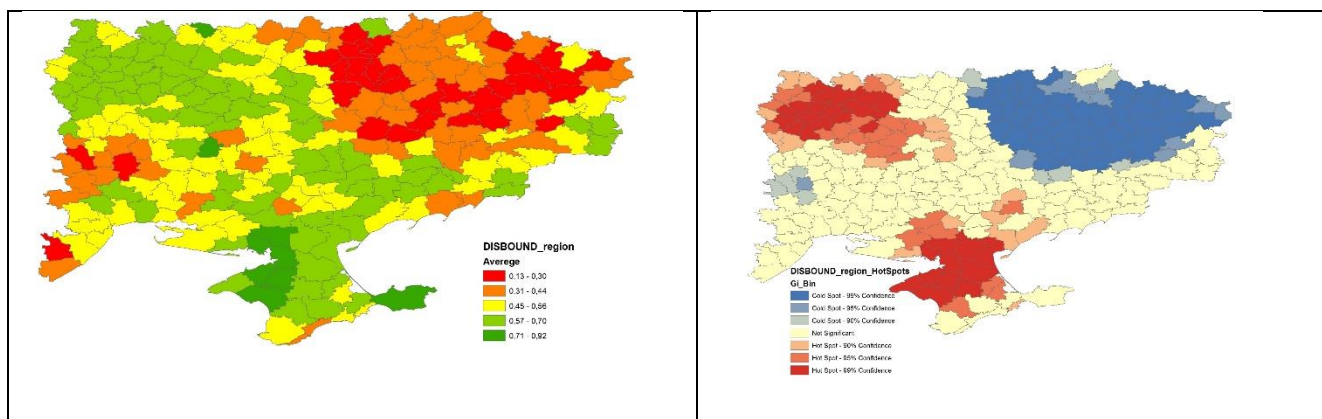


Рис. 3.13. Розподіл результуючого індексу за результатами класифікації а) та кластеризації методом «Гарячих точок» б).

Отримані статистичні кластери підтверджують гіпотезу про існування регіональних трендів погіршення вегетації у 2015 р. у порівнянні з 2014 р. у північно-східній частині України та дещо покращення цих показників для Херсонської області та АР Крим, а також Кіровоградської, Черкаської областей.

Висновки до розділу

Застосування методів класифікації та кластерного аналізу у поєднанні з аналітичними можливостями геоінформаційних систем значно розширює методичний інструментарій сучасного інтелектуального аналізу геоданих.

Метою класифікації є отримання деякої наперед заданої або не заданої кількості груп об'єктів (класів об'єктів). У межах кожного класу об'єкти повинні бути максимально «схожі» один на одного в деякому сенсі, однорідні, а об'єкти з різних класів - максимально «відрізнятися».

Визначено основні підходи до кластеризації геопросторових даних у випадках використання ієрархічних та неієрархічних методів кластеризації. Встановлено можливості організації багатовимірного простору з урахуванням геопросторового положення в двовимірному та тривимірному просторах. В цьому випадку можна виділити варіанти сумісного використання ознак геопросторового положення та непросторових ознак, стадійного застосування перед початком основної процедури кластеризації та після та незалежної кластеризації в різних ансамблях методів.

Методи класифікації без навчання поділено на два великі класи алгоритмів:

- методи, засновані на описі класів ядрами (паралельні і послідовні);
- ієрархічні процедури (агломеративні і дивізійні).

Виділено оціночні і типологічні класифікації.

$$S_1 < S_2 < \dots < S_K \text{ і } S_1 > S_2 > \dots > S_K$$

Під позначенням $S_1 < S_2$ розуміється, що будь-яка ОТО і-го класу краще ОТО j-го класу за комплексом показників. Вважається також, що в межах кожного класу ОТО приблизно однаково хороші або однаково погані (в розрізі проблеми, описуваної показниками).

Зазначено, що розробка методів класифікації, які дозволяють враховувати при проведенні аналізу географічний простір, ведеться в основному в руслі робіт по районуванню. Запропоновано, що модифікація існуючих методів автоматичної класифікації для цілей районування може здійснюватися таким чином:

- шляхом зміни формул заходів близькості між об'єктами за рахунок відображення в них відносного географічного положення об'єктів, що дозволяє врахувати топологічні властивості об'єктів в заходи подібності;
- шляхом зміни формул функціоналу якості розбиття на класи, які повинні бути за змістом ширше стандартних за рахунок включення територіальної складової;
- шляхом модифікації алгоритмів автоматичної класифікації, зазвичай за рахунок включення додаткових кроків перевірки географічної суміжності об'єктів.

Показано, що методи районування відповідають одному з підрозділів прикладної статистики, який носить назву класифікації при заданих обмеженнях на зв'язку між об'єктами. Наведені приклади не географічних постановок задач класифікації, в яких могли б використовуватися методи районування (і навпаки, методи яких могли б використовуватися для районування). На основі канонічної системи методів класифікації сформована система методів районування. Розібрані алгоритми деяких найбільш застосовних на практиці методів районування, які є модифікаціями канонічних методів, і метод бар'єрів максимальних відмінностей. Розроблено алгоритм побудови моделі сукупності однорідних територіальних зон.

Запропоновані засоби проведення кластерного аналізу розроблено у програмному середовищі ArcGis, що дозволяють класифікувати просторові об'єкти за просторово-атрибутивними параметрами. До основних переваг багатовимірного аналізу в середовищі ГІС слід віднести автоматизоване картографування кластерів, використання додаткової географічної інформації, варіабельність у прийнятті рішень щодо аналізу. Ієрархічні модифіковані для аналізу геопросторових даних методи було застосовано для аналізу розподілу відходів у регіонах України.

В результаті апробації програмного модуля кластеризації при дослідженні даних про поводження з відходами було зроблено декілька висновків:

У порівнянні з кластеризацією за атрибутивними параметрами застосування методу із просторово-атрибутивними параметрами надало можливість виділення більш рівномірних за кількістю адміністративних одиниць та більш територіально зв'язаних кластерів.

Отримані результати дослідження доповнюють результати зарубіжних досліджень просторової кластеризації, тому що дозволяють використовувати разом із ієрархічною моделлю кластеризації топологічні моделі, засновані на координатному просторі, що дозволяє виділяти кластери із врахуванням умов сусідства під час однієї процедури обробки та оцінювати взаємний вплив існуючих небезпек.

Серед 7 виділених кластерів перші 4 відносять-ся до регіонів з відносно високим рівнем небезпеки виникнення НС, 3 наступних кластери мають помірну небезпеку, пов'язану з поводженням з відходами.

Неієрархічні методи було застосовано для апробації розробленого підходу на основі кластерного аналізу даних про параметри сталого розвитку в Україні. На основі застосування просторово модифікованих методів k-середніх та нечітких k-середніх визначено на основі кластеризації у тривимірному та п'ятивимірному просторі основні кластери регіонів України за складовими показниками сталого розвитку з врахуванням геопросторового положення.

Система збору, накопичення та обробки супутникових даних Copernicus надає цінну інформацію про стан вегетації на планеті, яка може бути використана

у сільському господарстві, досліджені клімату та землекористування. Вивчення динамічних рядів цих даних дозволяє визначити довгострокові просторові тренди у зміні кліматичних умов та стану рослинного покриву. Сучасний математичний апарат системного аналізу та інструментарій геоінформаційних систем дозволяють отримати інтегральну оцінку цих просторово-часових трендів із використанням методів сумісного аналізу даних різної природи. Застосування цього підходу до даних про стан вегетації в Україні з використанням чотирьох видів індексів дозволило визначити значні регіональні диспропорції у динаміці рослинного покриву на основі кластерного підходу.

РОЗДІЛ 4 ГЕОСТАТИСТИЧНЕ МОДЕЛЮВАННЯ В АНАЛІЗІ ДАНИХ СТАЛОГО РОЗВИТКУ

4.1. Проблеми просторового моделювання

У цьому розділі ми розглянемо постановку задачі при аналізі і моделюванні просторових даних і наведемо приклади типових задач.

У 1986 році відбувся викид радіоактивних речовин з реактора на Чорнобильській АЕС. Радіоактивне забруднення розповсюдилося по повітрю на сотні кілометрів і торкнулося багатьох європейських країн [238]. Вимірювання радіоактивного забруднення ґрунту проводилися в багатьох місцях. Постали питання: «Як побудувати карту забруднення? Чи можна обійтися простими методами інтерполяції?» Чи можна дати однозначну відповідь про те, де проходить межа підвищеного рівня забруднення? На ці та багато інших питань можуть дати відповідь аналіз і моделювання просторових даних з використанням статистичних методів [75,76].

Існує величезна кількість просторово розподіленої інформації, зібраної в бази і банки даних з навколишнього середовища. Завдання її інтерпретації, аналізу та подальшого використання є надзвичайно важливою і вимагає комплексного системного підходу. Статистичне моделювання просторових явищ дозволяє узагальнити наявні вимірювання і отримати модель їх розподілу в просторі [32].

Найбільш поширеною проблемою при роботі з просторово розподіленими даними є отримання просторової оцінки. Так, було підготовлено багато різних карт по радіоактивному забрудненню ґрунту в результаті Чорнобильської аварії [179]. При цьому залишалося відкритим питання про якість і точності цих карт, невизначеності оцінки, чутливості використаних методів інтерполяції і т. п.

Таким чином, просторове моделювання є основою побудови моделей в різних сферах людської діяльності. Так, при кліматичному моделюванні аналізуються вимірювання температури, опадів, швидкості вітру і т. д. в різних точках простору. При моделюванні забруднення навколишнього середовища використовуються вимірювання (проби ґрунту, води, повітря, дистанційне зондування) в різних

місцях. У завданнях геології моделюються властивості порід в проміжку між свердловинами, де робляться вимірювання. У медичній географії аналізуються фактори, що впливають на рівень захворювань, і моделюється поширення епідемій. Просторово розподілені дані використовуються при моделюванні запасів корисних копалин, рибних ресурсів, криміногенної ситуації, природних катастроф (зсувів, лавин та ін.). [165]

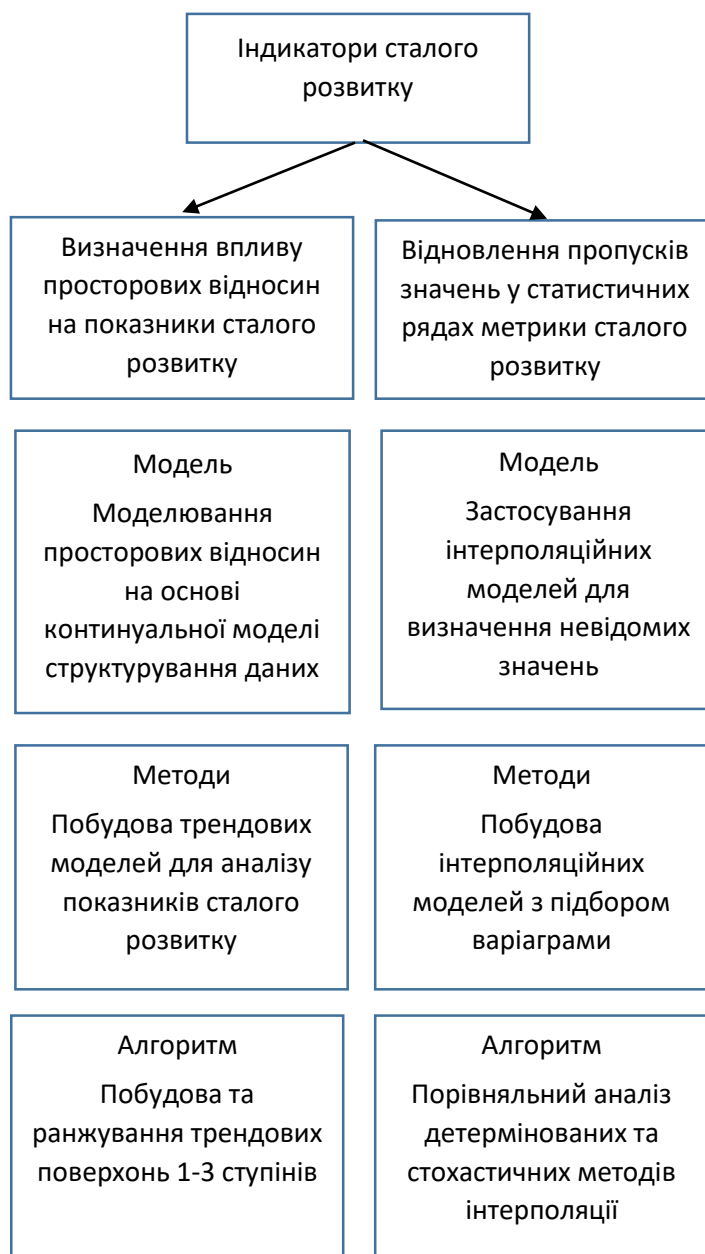


Рис.4.1. Модель використання методів геостатистичного аналізу для даних сталого розвитку

Глибокий аналіз та моделювання просторових даних вимагають застосування комплексного підходу і різних методів, що характеризують ту чи іншу особливість явища. Складність такого аналізу обумовлена декількома факторами: наявністю великих обсягів кількісної та якісної інформації по досліджуваному явищу, великим масштабом і мінливістю, наявністю різних факторів впливу [165].

Опишемо методологію геостатистичного аналізу просторово розподілених даних для дослідження складових сталого розвитку. Можна виділити дві складові геостатистичного аналізу процесу сталого розвитку: дослідження просторового розподілу методами геостатистики складових індексу сталого розвитку та визначення відсутніх значень методами інтреполяції для вхідних показників інтегральних індексів. Поєднання цих двох завдань дозволяє перейти до континуального розподілу складових та деталізації моделей сталого розвитку.

Постановка задачі

При роботі з просторовими даними зазвичай є певна кількість вимірювань досліджуваної змінної в різних точках, число яких обмежено. Таким чином, є область, на якій проведено ряд вимірювань деякої величини Z . Ці виміри проведено в, доволіно поширеному по області, наборі точок (x, y) , які ми будемо називати мережею моніторингу (рис. 4.2). Але в цій області є ділянки області, які не покриті вимірами, в яких потрібно отримати інформацію про значення величини Z . Як правило, потрібно розв'язати задачу інтреполяції, яка полягає в оцінці значення в невимірюваній точці на основі наявних даних.

В більшості випадків дані вимірювань дискретні і просторово неоднорідно розподілені. Таким чином, аналіз даних і його результати суттєво залежать від якості і кількості вихідних даних, від методів і моделей обробки даних.

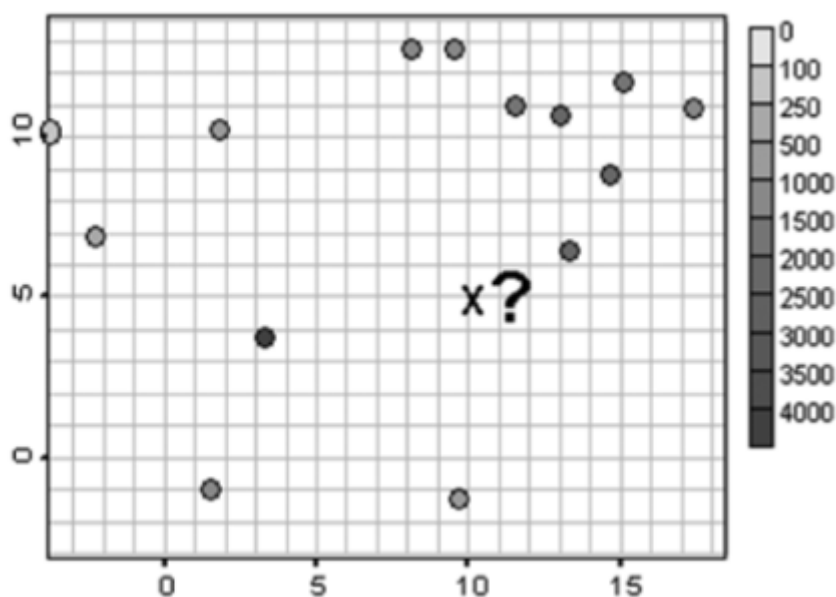


Рис.4.2. Постановка задачі просторової оцінки

При розв'язанні проблеми сталого розвитку з врахування безпеки життєдіяльності виникають комплексні задачі, для вирішення яких необхідне застосування комплексу спеціальних досліджень прихованих взаємозалежностей та взаємозв'язків, які ми можемо виявити лише за допомогою методів геостатистики – статистики просторово розподіленої (регіональної) інформації:

- побудувати та оцінити нові індикатори сталого розвитку в точках, де відсутні вимірювання;
- результати розрахунків цих індикаторів сталого розвитку візуалізувати в порівнюваних тривимірних форматах карт, побудувати ізолінії (визначити значення на щільній сітці);
- оцінити похибки інтерполяційних оцінок індикаторів;
- оцінити значення змінних з малим числом вимірювань, використовуючи значення інших корелюючих з ними змінних, з достатнім числом вимірів;
- знайти ймовірності перевищень значень спостережуваних змінних в потрібних областях;
- отримати набір єдино-можливих стохастичних просторових реалізацій розподілу спостережуваної змінної.

Таким чином, перші три задачі вимагають вдосконалення та побудови методів для просторової регресії. (для запропонованого набору індикаторів).

Дві останні задачі відносяться до проблеми знаходження та оцінювання імовірностей значень нових індикаторів, що пов'язані з ситуаціями, які характеризують моменти вимірювань і пов'язані з оцінками ризику.

Підходи до аналізу просторового розподілення даних

Існує кілька підходів до аналізу та обробки просторово розподілених даних, які можна умовно розділити на три групи:

Детерміновані моделі (інтерполятори) - лінійна інтерполяція на основі тріангуляції, метод зворотних відстаней в степені, мульти-квадратичні рівняння і т. п. [75];	Геостатистика - моделі, що базуються на статистичній інтерпретації даних [189];	Алгоритми, засновані на навчанні - штучні нейронні мережі, генетичні алгоритми, статистична теорія навчання машин векторів підтримки (Support Vector Machines) [284].
--	---	---

Звичайно, цей поділ до певної міри умовно. Так, геостатистичні моделі можна викласти в детермінованому формулюванні, і навпаки, деякі детерміновані моделі мають близькі статистичні аналоги. У свою чергу, статистичний підхід, на якому базується геостатистика, включає регресивні моделі просторових інтерполяцій (передбачень) і методи стохастичного моделювання. Алгоритми, засновані на навчанні (або штучний інтелект), також мають статистичну інтерпретацію.

Сучасна геостатистика – це широкий спектр статистичних моделей та інструментів для аналізу, обробки та подання просторово розподіленої інформації [234]. Нижче ми докладно опишемо найбільш часто використовувані моделі та інструменти, з яких можна скласти замкнений цикл дослідження і вирішити поставлені вище завдання.

Традиційні детерміновані методи широко використовуються для просторової інтерполяції і дозволяють вирішувати тільки першу і другу задачі з наведеного

вище списку. Геостатистична теорія дозволяє вирішувати весь набір завдань, в тому числі оцінити невизначеність оцінки і описати її варіабельність.

Геостатистика виникла на початку 1960-х рр. як теорія регіональних змінних, сформульована Ж. Матерон (Matheron) для аналізу даних про природні копалини (гірничорудна справа) [96]. Він організував Центр геостатистики в Фонтенбло. Цей центр вніс помітний вклад у теоретичні дослідження і їх практичні застосування.

Незалежно від Ж. Матерон і практично в той же час Л. С. Гандіні сформулював теорію оптимальної інтерполяції для об'єктивного аналізу метеополей [25]. У цій теорії також наведені основи геостатистичної теорії. На жаль, наступні роботи російських вчених в цій області не знайшли в той час широкої підтримки [20].

Сучасна геостатистика - це область прикладної статистики, що швидко розвивається, з величезним набором лінійних і нелінійних методів, параметричних і непараметричних моделей для аналізу, обробки та подання просторово-часової інформації. Спектр її застосування досить широкий - від традиційного використання в області видобутку копалин до сучасних додатків в економіці, фінансах, навколишньому середовищу, епідеміології і т.п. [264].

Геостатистичний аналіз дозволяє створити СППР нового типу значно підвищити рівень надійності та якість рішень, що приймаються саме на основі використання просторово розподіленої інформації. Сучасні тенденції геостатистики пов'язані з розвитком методів стохастичного моделювання (просторових аналогів методів Монте-Карло), методів, заснованих на багатоточковій статистиці, гібридних моделях з широким використанням алгоритмів штучного інтелекту та додаткової інформації різного виду і додатками в області обробки і передачі зображень, з розширенням на часовий і просторово-часовий аналізи по багатьох актуальних напрямках [288].

Відмітимо, що важливою складовою традиційної та новітньої геостатистики є просторовий кореляційний аналіз або варіографія [134]. На основі розробленого методу [121] варіографія дозволяє провести структурний аналіз та побудувати адекватні моделі, що дають можливість виявити приховані залежності нових явищ та процесів соціально-економічної сфери та їх тенденцій, а також передбачити

системний розвиток [136]. Крім того це дозволяє зробити глибокі висновки про статистичну природу даних і структурі адекватних моделей. В принципі експериментальна варіографія, заснована на вихідних даних, може бути використана в більшості завдань просторового оцінювання незалежно від методу інтерполяції нарівні з традиційним статистичним аналізом. [124].

Основні етапи аналізу та моделювання просторових

1. Збір та підготовка даних до обробки.
2. Другим і дуже важливим етапом дослідження є сучасний статистичний аналіз просторо-часових даних, що дозволяє визначити наявність помилок і викидів (outliers) в даних, оцінити базові статистичні закономірності, провести кореляційний аналіз при наявності декількох змінних і т. п.
3. Розробка моделі сталого розвитку з врахуванням безпекової компоненти і просторових даних та відповідного математико-статистичного забезпечення.

При моделюванні сталого розвитку з безпековою складовою виникає так звана просторова кластеризація, яка має нерегулярний характер просторового розподілу в залежності від щільності проведених вимірів, що призводить до втрати репрезентативності статистичних даних – середніх, варіацій, гістограм. Якщо мережа моніторингу має зони з помітно більш високою щільністю вимірювань, ніж решта область, то мережа моніторингу кластерна. При цьому зони підвищеної щільності вимірювань характеризуються більш високими (або, навпаки, низькими) значеннями вимірювань, і тому виникає необхідність в декластеризації. Інакше оцінки всіх статистичних характеристик будуть спотворені, наприклад оцінка середнього буде завищена (або, навпаки, занижена). Процедура декластеризації орієнтована на усунення такого роду спотворень. Виділяють два основних типи декластеризації - вибірккову і зважену. Вибіркова декластеризація пов'язана з вибором частини даних з кластерів, Зважена передбачає завдання ваг, які враховують щільність вимірювань.

Оцінити деякі просторові особливості даних дозволяє статистика з рухомим вікном: область розбивається на під-області, в кожній з яких проводиться незалежний статистичний аналіз [165].

Подальший просторовий аналіз передбачає дослідження і моделювання просторової кореляції між даними по одній або декільком змінним. Мірою просторової кореляції є варіограма – статистичний момент другого порядку.

Для отримання «найкращої», в статистичному сенсі, просторової оцінки використовуються моделі з сімейства крігінга (kriging) - найкращого лінійного незміщеного оцінювача (best linear unbiased estimator - BLUE).

Крігінг є оптимальним оцінювачем в статистичному сенсі в класі лінійних інтерполяторів - його оцінка є мінімальною варіацією помилки. Важлива властивість крігінга - точне відтворення значень вимірів в наявних точках (точний оцінювач). На відміну від численних детермінованих методів, оцінка крігінга супроводжується оцінкою помилки інтерполяції в кожній точці. Отримана помилка дозволяє охарактеризувати невизначеність отриманої оцінки даних за допомогою довірчих інтервалів.

При застосуванні будь-якої моделі інтерполяції постає питання про підбір оптимальних модельно-залежних параметрів. Можна бачити, що в разі використання одного і того ж методу інтерполяції можна отримати якісно різні результати в залежності від вибору модельних параметрів. Вибір оптимальних параметрів спирається на покрокове дослідження характеру і структури даних. Ефективними інструментами підбору модельних параметрів є методи крос-валідації (crossvalidation), складного ножа (jack-knife), бутстреп (bootstrap) [199]. Всі вони засновані на проведенні оцінки для частини точок вимірювань, обраних з основного набору по іншим даними з подальшим обчисленням помилки оцінки. Після оцінок по всіх точках, наборах або вибірках оцінюється середньоквадратична помилка отриманих оцінок. По ній порівнюються різні методи або вибираються найкращі параметри методу. У геостатистиці традиційно більш широко використовується крос-валідація [32].

При проведенні аналізу реальних даних експерти часто стикаються з проблемою малої кількості вимірювань щодо необхідної змінної, наприклад внаслідок їх великої ціни або небезпеки взяття проб. При цьому в наявності може виявитися велика (надлишкова) кількість «дешевих» вимірювань змінної, яка досить сильно корелює з основною. Постає питання, як можна використовувати «дешеву» інформацію для поліпшення оцінки змінної, інформація по якій «дорога». В рамках багатозмінної геостатистики існує модель спільної просторової інтерполяції декількох корельованих змінних – кокрігінг. Кокрігінг дозволяє значно поліпшити якість оцінки, перейти з області екстраполяції в область інтерполяції, зменшити помилку оцінки за рахунок використання додаткової «дешевої» інформації по скорельованим змінним [320].

Часто результатом просторового аналізу даних в рамках кваліфікованої підтримки прийняття рішень є імовірнісні карти. Імовірнісне картування дає можливість оцінити рівень ризику по перевищенню або не перевищенню заданого рівня значення просторової змінної. Таке картування використовується при оптимізації рішень, коли просторовий аналіз даних є тільки проміжним етапом. В рамках геостатистики для імовірнісного картування використовуються нелінійні моделі крігінга, зокрема індикаторний крігінг, який дозволяє розрахувати локальну функцію розподілу в точці оцінювання. Карти ймовірностей середніх оцінок, карти оцінок із заданою вірогідністю перевищення використовуються в процесі прийняття рішень.

Застосування різних детермінованих або геостатистичних моделей інтерполяції / оцінювання завжди дає єдине і згладжене, яке не відтворює початкову варіабельність даних значення оцінки в точці, що цікавить, при обраних модельних параметрах. Стохастичне моделювання є альтернативним підходом, що дає можливість відтворити вихідну варіабельність і отримати скільки завгодно багато рівно-ймовірнісних реалізацій просторової функції в області. Рівно-ймовірні реалізації дають можливість окреслити просторову варіабельність (мінливість) і невизначеність просторової функції, оцінити ймовірності і ризик. При використанні стохастичного моделювання вдається уникнути «згладженої»

картини оцінки, яка властива більшості моделей інтерполяції. Це дозволяє отримувати коректні результати в таких задачах, як, наприклад, розрахунок обсягу нафтового резервуару, «довжини» берегової лінії і т. п. [32].

Наведемо блок-схему описаних етапів покрокового аналізу і моделювання просторових даних (рис. 4.2). В її основу покладена методологія, яка випробувана в численних дослідженнях сталого розвитку, в тому числі і з врахуванням даних радіоактивного Чорнобильського забруднення [76]. Слідуючи розробленій методології докладно опишемо спектр алгоритмів, які застосовуються на кожному етапі СППР.

Для виявлення прихованих нелінійних зв'язків та побудови нелінійних моделей в СППР включені блок застосування штучних нейронних мереж для використання геопросторової інформації при вирішенні завдань просторової регресії і класифікації

Статистичні моделі котрі навчають, такі як штучні нейронні мережі і машини підтримуючих векторів (support vector machines), можна використовувати поряд з геостатистичними моделями для вирішення завдань просторової регресії і класифікації [288].

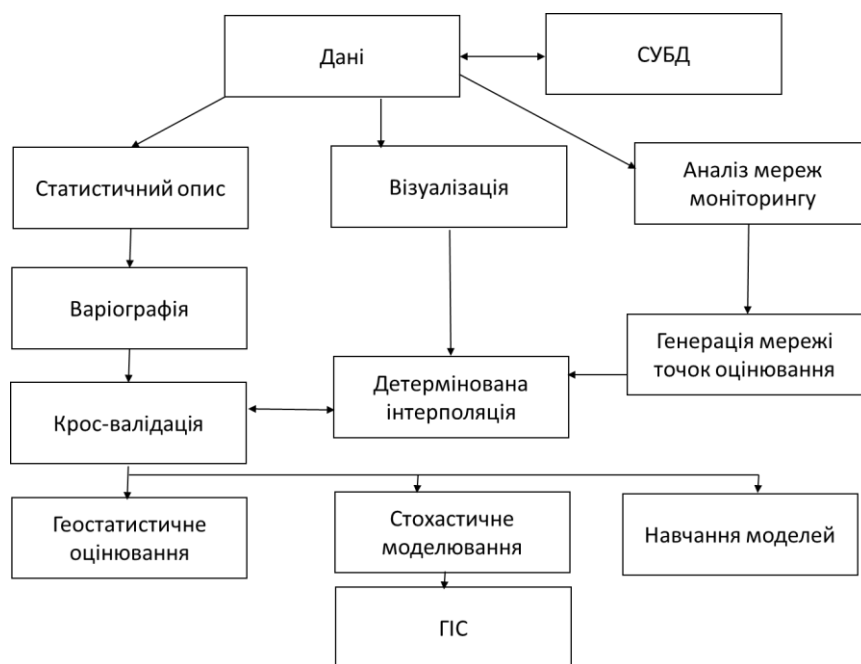


Рис. 4.3. Блок-схема методології послідовного аналізу та моделювання просторово-розподілених даних

4.2 Основні поняття та елементи геостатистики

Цей підрозділ присвячено базовим поняттям і припущенням геостатистики, а також суміжних областей. Дано визначення просторово розподілених даних, з якими працює геостатистика.

Просторово розподілені дані

При аналізі даних різних вимірів часто вкрай складно або зовсім неможливо отримати формульний закон розподілу даних на основі фізичних процесів, які обумовлюють відповідні явища. Альтернативний підхід - статистичний (а не детермінований) опис просторового розподілу. На відміну від детермінованих методів геостатистичні оцінки спираються на інформацію про внутрішню структуру даних, залежать від самих даних, тобто є адаптивними. Геостатистика базується на статистичній інтерпретації даних. Передбачається, що дані вимірювань $z(x_i)$ є реалізаціями випадкових змінних $Z(x_i)$, які описуються певними функціями розподілу. Це, однак, не означає, що природа самого процесу є випадковою. Щоб використовувати геостатистику, необхідно визначити просторову кореляційну структуру поля $Z(x)$, що задається усіма випадковими змінними в області дослідження. Геостатистичний підхід дозволяє виходити при аналізі з строгих критеріїв.

Предметом аналізу геостатистики є *просторові змінні* (або регіоналізовані змінні - regionalised variables), які аналогічні змінним з координатною прив'язкою. Приклади просторових змінних: кількість опадів, щільність населення в деякій географічній області, потужність геологічної формації, щільність забруднення ґрунту, середнє споживання електроенергії в певний час і т. д. Просторові змінні не слід плутати з випадковими величинами, які досліджуються методами загальної статистики.

Випадкова функція визначається як набір залежних між собою випадкових змінних $Z(x_i)$, по одній для кожного місцезнаходження x_i в розглянутій області. Будь-якому набору з N місцезнаходження $\{x_k, k = 1, \dots, N\}$ можна поставити у

відповідність N випадкових змінних $\{Z(x_1), \dots, Z(x_N)\}$, які характеризуються N -мірною умовної функцією розподілу:

$$F(x_1, \dots, x_N; z_1, \dots, z_N) = \text{Prob}\{Z(x_1) \leq z_1, \dots, Z(x_N) \leq z_N\}. \quad (4.1)$$

Нагадаємо, що поняття *випадкової величини* в класичній статистиці має конкретний сенс тільки при дотриманні наступних умов:

1) повинна бути хоча б теоретична можливість нескінченного повторення випробувань (реалізацій), в результаті яких випадкова величина набуває чисельні значення;

2) результат кожного з випробувань повинен бути незалежним від результатів всіх попередніх випробувань.

Просторова змінна не задовольняє жодної з цих умов. Якщо випробування полягає у відборі проби в точці x , то вміст шуканої речовини в такій пробі буде єдиним, фізично визначеним і ні в якій мірі не випадковим. Оскільки немає ніякої можливості повторити вказане випробування, то проба в конкретній точці не підлягає виконанню умови 1. Однак є можливість відібрати нову пробу в околі точки x , тому можна наближено забезпечити виконання умови 1. Але тоді порушується умова 2: якщо перша проба відібрана зоні значимій точці, то друга проба, узята в безпосередній близькості від першої, як правило, теж буде значима. Таким чином, випробування виявляються залежними.

Надалі для зручності будемо використовувати для зручності звичний в статистиці термін *випадкової величини*, розуміючи під нею просторову регіоналізовану змінну з врахуванням вищеописаних особливостей.

Досліджувана змінна може бути неперервною (наприклад, будь-яка фізична величина - щільність, тиск, концентрація і т. д.) або категоріальною (наприклад, тимчасовий бінарний сигнал або тип ґрунту або геологічної породи). Тому для аналізу змінних різного типу природньо використовувати різні підходи.

Метрика в просторі

Ми будемо розглядати так звані регіоналізовані дані, а саме вимірювання, що мають координатну прив'язку. Координатна прив'язка може бути:

- просторовою, що визначає географічне положення вимірювання (географічні координати) в просторі або його відносне положення по відношенню до інших об'єктів (спеціальна координатна система для певної місцевості);
- часовою, що визначає час проведення вимірювання (абсолютне або відносне);
- просторово-часовою, тобто і просторовою, і часовою одночасно.

Таким чином, основна вимога до координатної системи – її метричність, тобто координати повинні супроводжуватися метрикою, можливістю обчислювати відстані між точками. Надалі, якщо інше не зазначене, будемо вважати, що точки простору належать R^2 , де метрика – це відстань між точками простору $X_1 = (x_1, y_1)$ і $X_2 = (x_2, y_2)$ і визначається евклідовою нормою:

$$\|(x_1, y_1)(x_2, y_2)\| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (4.2)$$

Введення третьої та інших просторових координат призводить до евклідової відстані між точками $X_1 = (x_{11}, \dots, x_{1n})$ і $X_2 = (x_{21}, \dots, x_{2n})$ в n -вимірному просторі та обчислюється аналогічно двовимірному випадку:

$$\|(\bar{X}_1, \bar{X}_2)\| = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}. \quad (4.3)$$

Потрібно відмітити, що слід враховувати різницю масштабів вимірювання компонент складових вектора. Наприклад, якщо розглядати поширення забруднення в ґрунті, то горизонтальні просторові області поширюються на десятки кілометрів (10^4 - 10^5 м), в той час як досліджувана глибина при аналізі ґрунту не перевищує 0,3 м, а геологічні шари можуть мати товщину від декількох сантиметрів до сотень метрів. При роботі з такими різними масштабами зазвичай проводиться нормування координат - перехід до іншої системи, де розмірності зрівнянні, наприклад лінійне перетворення на відрізок (0, 1).

Просторова змінна завжди визначена в конкретній області простору - в *геометричному полі*. Просторову змінну V можна розглядати як функцію точки простору x : $Z = Z(x)$. Однак частіше інтерес представляють не точкові, а середні значення величини $Z(x)$ в межах малої області простору (носія) - *геометричній базі*. Наприклад, для такої ознаки, як вміст хімічного елементу в ґрунті, геометричною

базою є об'єм проби. База повинна бути визначена вельми точно. Необхідно знати її об'єм, форму і орієнтацію в просторі. Якщо змінюється геометрична база, то виникає нова просторова змінна, близька до попередньої, але яка не збігається з нею:

$$z^*(x_0) = \frac{1}{S} \int_{S(x_0)} Z(x) dx. \quad (4.4)$$

Таким чином, теорія просторових змінних дозволяє прогнозувати характеристики змінної Z , пов'язаної з геометричною базою в полі S , за відомими характеристиками іншої точкової змінної V , заданої в полі X , відмінному від поля S .

Просторова роздільна здатність

Однією з ключових властивостей просторово розподілених даних є їх просторова роздільна здатність. При дослідженні того чи іншого просторового явища важливо, щоб наявні дані могли адекватно його відобразити. При цьому під просторовою роздільною здатністю мається на увазі найменший розмір особливості, яку можуть відображати дані і просторові оцінки.

Розрізнення інтерполяційної просторової оцінки на регулярній сітці характеризується розміром комірки. Якщо сітка оцінювання нерегулярна, то її роздільну здатність можна охарактеризувати розподілом відстаней між вузлами сітки (див. нижче).

Ефект роздільної здатності сітки оцінювання може бути значним, особливо при вирішенні динамічних завдань з граничними умовами на сітці. У статичних задачах просторового картографування роздільна здатність сітки також має велике значення. Надалі приймаємо характеристики [216], які пов'язують роздільну здатність сітки з картографічним масштабом.

Носії (support) даних вимірювань є однією з основних властивостей при аналізі просторово розподілених даних. Носій не слід плутати з просторовою роздільною здатністю моделі (карти інтерполяційної оцінки). Носій характеризується процесом вимірювання і обробки даних, а не моделюванням. Під носієм вимірювання розуміється фізичний об'єм, підданий виміру. Наприклад, при

вимірюванні радіоактивності зразка опора вимірювання характеризується розміром проби. Однак не завжди вдається однозначно оцінити опорний розмір: так, при аерогаммазйомці забруднених територій носій може варіюватися від десятків до сотень метрів.

Визначення розміру носія даних вимірювань, що використовуються в моделях просторового оцінювання, надзвичайно важливе для адекватного моделювання варіабельності даних. Так, якщо при інтерполяції на сітку з роздільною здатністю 1 км використовуються дані з опорою 10 см, треба розуміти, що такі дані мають варіабельність на підсітковому масштабі. Іншими словами, величина даної змінної в комірці сітки оцінювання не може бути однозначно визначена на основі даних з опорою дрібнішого масштабу.

При моделюванні властивостей пористості і проникності порід в підземних родовищах розміром кілька кілометрів використовуються дані з різними носіями. Так, пористість і проникність, виміряні на основі кернів з свердловин, мають високу точність і опору порядку декількох сантиметрів. Дані ж сейсмічного зондування мають зашумленість, і розмір їх опори не завжди вдається однозначно визначити (від одиниць до сотень метрів). Динамічні вимірювання тиску в свердловині мають опору порядку декількох кілометрів, оскільки відображають поведінку середовищ в пов'язаній пористій системі родовища. Все це необхідно враховувати при моделюванні невизначеності і варіабельності просторових розподілів на основі даних різних типів.

Мережа моніторингу та кластерність

Найпростішим загальноприйнятим видом візуалізації даних є нанесення точок на площину просторових координат, причому колір нанесеної точки може характеризувати виміряні в них величини (рис. 4.4а).

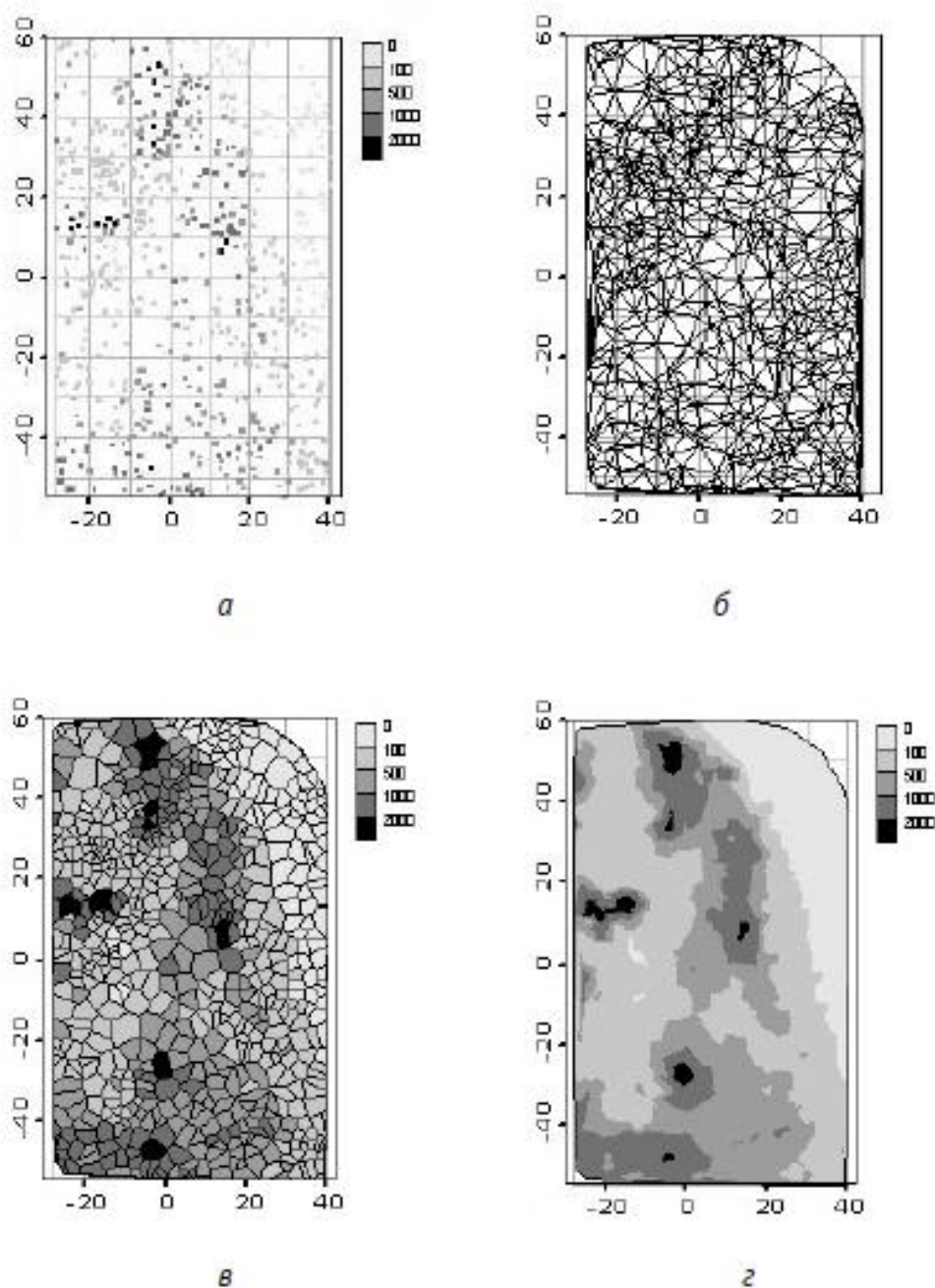


Рис.4.4. Діаграма розташування точок вимірювань (а), триангуляція мережі моніторингу (б), полігони Вороного (в) і контури даних вимірювань по триангуляції (г)

Для візуалізації мережі моніторингу та її кластерної структури часто використовується *триангуляція Делоне* [122] - система трикутників з вершинами в точках виміру, непересічними ребрами і мінімальною кількістю тупокутних трикутників (рис. 4.3б). Така візуалізація дозволяє якісно відокремити області з підвищеною щільністю вимірювань - з кластерами. Крім того, триангуляція Делоне

будує *систему сусідства*: точки, які з'єднані один з одним ребрами трикутників, є найближчими сусідами по відношенню один до одного.

Тріангуляція також є основою для побудови найпростішого методу лінійної інтерполяції: три точки в просторі (вершини трикутників) однозначно визначають площину, в межах якої значення функції обчислюються згідно геометричних принципів (рис. 4.4 г).

Іншим видом візуалізації даних є *полігони Вороного*, або, як їх ще називають, *розбиття Тіссена*, *комірки Діріхле* і *області впливу*. Полігон Вороного P_i , побудований для точки вимірювань x_i , характеризується тим, що містить ті і тільки ті точки, відстань від яких до точки x_i менша або дорівнює відстані до будь-якої іншої точки вимірювань x_j (рис. 4.4в). При побудові полігонів Вороного використовується система сусідства, отримана в процесі тріангуляції Делоне. Межі полігону Вороного P_i складаються з відрізків серединних перпендикулярів, проведених до сторін трикутників Делоне. Полігони Вороного можна використовувати як розривну інтерполяційну оцінку (оцінка по найближчому сусідові). Для цього кожній точці, що потрапила в полігон, надається значення, відповідне до його материнської точки. Ці полігони також використовуються в задачах просторової класифікації – класифікація по найближчому сусідові.

Для виявлення особливостей, а саме наявності кластерних структур або розрідження в мережі моніторингу (наборі точок вимірювань), проводять аналіз мережі моніторингу. Найпростішими методами такого аналізу можна вважати опис топології мережі за допомогою гістограми відстаней між точками (рис. 4.5а) і гістограми площ полігонів Вороного (рис. 4.5б). Гістограма в даному випадку - це графік числа будь-яких подій (числа пар або числа полігонів), що потрапили в будь-який інтервал значень.

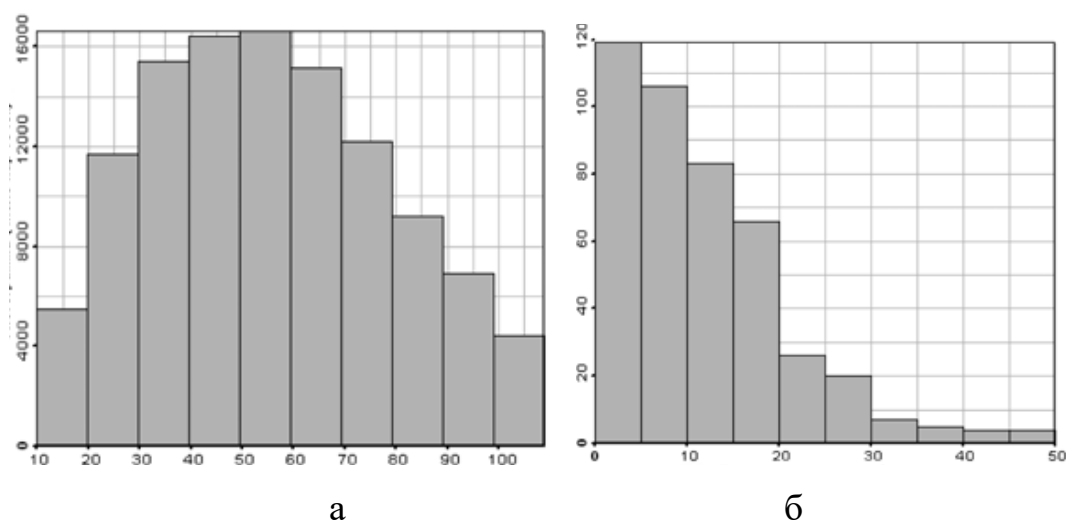


Рис.4.5. Гістограма відстаней між точками (а) і гістограма площ полігонів Вороного (б)

При рівномірному розподілі точок в просторі число пар повинно бути однаковим для всіх відстаней (або зменшуватися при збільшенні відстані за рахунок граничного ефекту). Зростання числа пар зі збільшенням відстані між точками свідчить про наявність кластерів. Гістограма площ полігонів для регулярної сітки повинна являти собою дельта-функцію (один пік), так як всі полігони одного розміру. Будь-які спотворення (широкий пік, довгий хвіст, кілька піків) означають присутність будь-яких особливостей в мережі.

Іншим методом аналізу мережі моніторингу є статистичний підхід [Cressie, 1991], який розглядає точки вимірювань як випадковий точковий процес. Характеризувати розподіл точок можна з використанням статистичних індексів. Прикладом такого підходу є *діаграма Морішіта*. Індекс Морішіта обчислюється для області, розбитої на прямокутні комірки рівного розміру, за формулою [159]

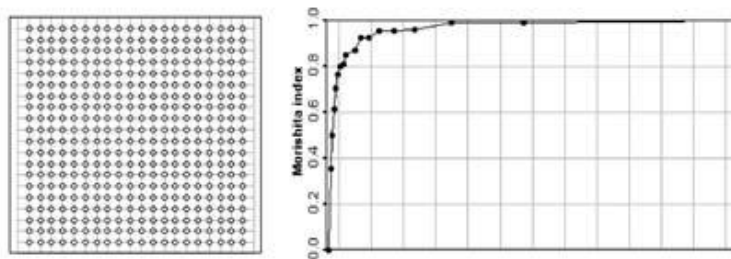
$$I_{\delta} = Q \frac{\sum_{i=1}^Q n_i(n_i-1)}{N(N-1)}, \quad (4.5)$$

де N - повне число точок мережі моніторингу; Q - кількість комірок розбиття; n_i ($i = 1, 2, \dots, Q$) - число точок мережі моніторингу, які потрапили в i -ю комірку. Цей індекс характеризує ймовірність того, що при виборі двох випадкових точок вони опиняться в одній комірці. Діаграма Морішіта являє собою залежність індексу Морішіта від розміру комірки розбивки. Існують три типи характерної поведінки

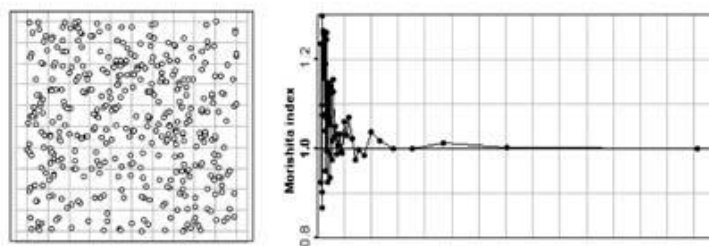
діаграми Морішіта, комбінації яких дозволяють судити про характеристики мережі моніторингу [129]:

- величина індексу Морішіта з ростом розміру комірки зростає і прямує до 1; тоді розподіл точок можна вважати рівномірним;
- величина індексу Морішіта не залежить від розміру комірки і приблизно дорівнює ≈ 1 (коливається близько 1); це означає, що розподіл точок випадковий і не має кластерних структур [129].
- величина індексу Морішіта з ростом розміру комірки зменшується або росте вище 1 - розподіл точок мережі кластерний.

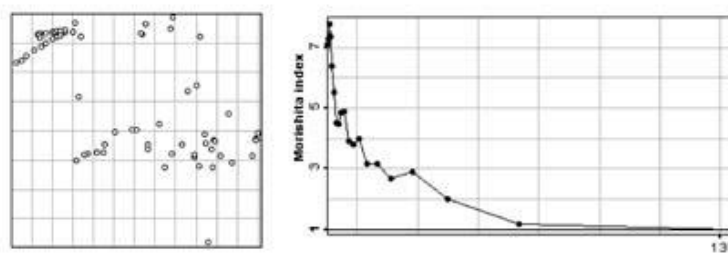
На рис. 4.6 наведені приклади діаграм Морішіта для різних типів мереж моніторингу. Так, у разі моніторингу на регулярній рівномірній сітці діаграма має вигляд гладкої кривої логарифмічного типу, яка прямує до одиниці (рис. 4.6а). При наявності численних кластерів в щільній мережі моніторингу крива Морішіта характеризується точками перегину, які визначають розміри різних кластерів (рис. 4.6б). У разі довільного моніторингу з декількома чітко вираженими кластерами крива Морішіта має більш гладкий вигляд і зменшується, прямуючи до одиниці (рис. 4.6в). Розмір кластерів характеризує в цьому випадку точки зміни кривизни.



а



б



В

Рис. 4.6. Приклади діаграми Морішіта для різних мереж моніторингу: регулярна рівномірна мережа (а), довільна мережа зі слабо кластерною структурою (б), довільна слабо зв'язана кластеризована мережа (в)

Для топологічно глибокого аналізу мереж моніторингу на неоднорідність потрібно також використовувати теорію фракталів і фрактальну розмірність [169] (характеристику ступеня самоподібності об'єкта). Методи обчислення і використання фрактальної розмірності детально розглянуті в [172].

Декластеризація

Більша частина просторово розподілених даних, які аналізуються в геостатистиці, мають кластерну структуру. Кластер утворюється, якщо в одній області було проведено значно більше число вимірювань, ніж в іншій. В цьому випадку виникають істотні спотворення при обчисленні, наприклад середнього значення. Наслідком цього є неможливість отримати репрезентативну гістограму розподілу. Наприклад в області високих значень вимірюваної величини, знаходиться вдвічі більше точок, ніж в області низьких значень. Якщо при оцінці середнього і інших статистичних параметрів всі значення матимуть однакову вагу, то область високих значень буде занадто сильно впливати на таку оцінку. В цьому випадку точки з зони з великими значеннями потрібно було б враховувати з вагою, вдвічі меншою, ніж всі інші. Проблема обчислення статистичної ваги кожної точки в параметрах розподілу вирішується шляхом проведення процедури *декластеризації* (declustering) даних.

Декластеризація не потрібна, якщо вимірювання були виконані на регулярній сітці. У цьому випадку опис розподілу найкраще вийде з рівною вагою. Крім того у багатьох випадках неможливо або небажано отримати дані на рівномірній сітці.

Таким чином, при аналізі вимірювань, проведених на нерегулярній сітці, передбачається існування такого набору ваг, при якому може бути отримано репрезентативний розподіл даних. При цьому дані з області з більшою щільністю вимірювань потрібно брати з меншою вагою (для зменшення їх впливу на розподіл в цілому), ніж дані з області з меншою щільністю вимірювань. Інші підходи для обчислення ваг можуть бути використані різні підходи: метод коміркової декластеризації, метод комірок Діріхле, крігінг.

Геостатистичне оцінювання

Основною геостатистичною моделлю, яка в тому чи іншому вигляді використовується в усіх методах геостатистики, є крігінг (kriging) - лінійний інтерполятор, що використовує для отримання оцінки значення функції в деякій точці простору x_0 експериментально виміряні значення цієї функції в інших точках:

$$Z^*(x_0) = \sum_{i=1}^{N(x)} w_i(x_0) Z(x_i). \quad (4.6)$$

Для визначення ваги $w_i(x_0)$ можуть бути використані різні детерміновані методи, наприклад ваги можуть братися обернено пропорційними відстані від виміряної точки до оцінюваної або відповідно до будь-якого іншого припущення про природу зв'язків в даних. Однак всі ці методи нехтують використанням інформації про структуру внутрішньої кореляції просторових даних.

Наступним критерієм при побудові моделі є умова *незміщеності оцінки*, що еквівалентно умові

$$E\{Z^*(x_0) - Z(x_0)\} = 0, \quad (4.7)$$

де $Z(x_0)$ - справжнє (невідоме) значення оцінюваної функції в точці x_0 . Іншими словами, помилки інтерполяції повинні мати в кожній точці середнє, рівне нулю. Ця умова може бути реалізована і в рамках детермінованих підходів.

Накладемо умови оптимальності інтерполяції в сенсі *мінімізації варіації помилки оцінки*, тобто ваги w_i лінійної регресії в рівнянні (4.6) повинні бути обрані так, щоб мінімізувати значення варіації помилки оцінки:

$$\text{Var}[Z^*(x_0) - Z(x_0)] = E\{[Z^*(x_0) - Z(x_0)]^2\}. \quad (4.8)$$

Таким чином, крігінг є найкращим (в сенсі мінімуму варіації оцінки) лінійним і незміщеним оцінювачем (the best linear unbiased estimator - BLUE). При цьому в процесі пошуку мінімуму варіації (4.8) ключову роль відіграє використання моделі варіограми вихідних даних.

Таким чином значення з варіації (4.8) чисельно визначає точність крігінгової оцінки.

Перевірка якості моделі – крос-валідація

При використанні тієї чи іншої моделі інтерполяції важливо правильно підібрати значення модельно-залежних параметрів. Для крігінга такими параметрами є параметри моделі варіограми. При роботі з реальними даними не завжди вдається відразу вибрати теоретичну модель експериментальної варіограми. Для перевірки якості обраної моделі використовують різні кількісні методи: крос-валідацію (cross-validation), метод складного ножа (jack-knife), бутстреп (bootstrap) [1].

Крос-валідація - найбільш простий і часто використовуваний не тільки в геостатистиці підхід при порівнянні результатів, отриманих різними методами або одним і тим же методом, але з різними параметрами. Виконується крос-валідація наступним чином:

- з бази даних тимчасово вибуває одна точка, і для неї проводиться оцінка значення;
- отримане значення порівнюється з відомим, і обчислюється нев'язка - різниця між вимірами і оціненими значеннями:

$$\Delta Z(x) = Z(x) - Z^*(x);$$

- перші два кроки проводяться для всіх точок бази даних.

Отримані нев'язки $\Delta Z(x)$ можуть бути графічно представлені у вигляді карти (карти нев'язок), по якій можна подивитися, в яких зонах метод спрацьовує краще, а в яких гірше. Замість нев'язок можна візуалізувати *відносні помилки*:

$$\text{relative error } (x_i) = \frac{Z(x_i) - Z^*(x_i)}{Z(x_i)}.$$

Корисно також представити результати крос-валідації у вигляді графіка $Y(Z(x)) = Z^*(x)$ або аналогічному йому - $Y(Z(x)) = \Delta(x)$. Проведення на такому графіку бісектриси (або відповідно прямої $Y = 0$), відповідної рівності оцінки і вихідного значення, дозволяє побачити характер відхилення: більше відхилення для високих або низьких значень Z , будь-які тренди в поведінці оцінки і т. д. Разом з тим на графіках нев'язок простежується ефект згладжування – область низьких значень в середньому переоцінюється, а область високих значень недооцінюється.

Крім локальних характеристик крос-валідація дозволяє оцінити і глобальні характеристики оцінки для порівняння:

1. Зсув $\Delta m = m - m^*$, де m - середнє, оцінене по вихідним даним; m^* - середнє, оцінене за отриманими результатами.

2. Суму квадратів нев'язок:

$$S = \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2 + R,$$

де R - штрафний член, який вводиться для контролю кількості неоцінених точок.

3. Середню квадратичну помилку (root mean square error - RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2}.$$

4. Коефіцієнт ефективності:

$$E = \frac{S}{S_0},$$

$$\text{де } S_0 = \sum_{i=1}^n \{Z(x_i) - m\}^2.$$

5. Коефіцієнт кореляції r , кут нахилу регресійної прямої на графіку $Y(Z(x)) = Z^*(x)$.

Взагалі говорячи, крос-валідація - це окремий випадок методу складеного ножа, коли обраний набір складається з однієї точки (leave-one-out). *Метод складеного ножа* (jack-knife) є загальним випадком крос-валідації, коли оцінювання проводиться не в одній, а в декількох точках вимірювань, дані про які попередньо вилучаються з розгляду. Отримані в результаті нев'язки аналізуються

методом, аналогічним описаному вище. Оскільки при jack-knife вилучається довільний набір даних, комбінації цього набору можуть варіюватися, що робить цей метод стохастичним.

Бутстреп (bootstrap) полягає в оцінці на основі випадкових вибірок з набору даних. Вибірки робляться з вихідного набору випадковим чином. Обрана точка не вилучається, вона може потрапити до вибірки кілька разів. Оцінка проводиться за залишеними не обраними точками. Зазвичай процедура вибірки і оцінки повторюється багато разів.

4.3 Детерміновані методи просторової інтерполяції

Детерміновані методи традиційно широко використовуються в різних областях прикладної і наукової діяльності. Наприклад, широко відомий пакет SURFER містить досить велику колекцію таких методів [217]. Наведемо деякі алгоритми, які найбільш часто зустрічаються, і відзначимо особливості їх використання. У трьох розділах цієї глави описані три основні підходи до детермінованої інтерполяції: лінійні моделі, поліноміальні моделі та моделі базисних функцій.

Детерміновані методи інтерполяції припускають наявність заданої аналітичної залежності між значеннями функції в просторі. Ці методи популярні через простоту обчислення оцінки за заданою параметричною формулою. Найбільш широко застосовуються «формульні» залежності: обернена пропорційність відстані (або його ступеня), сплайни, поліноми різних ступенів та ін. Однак детерміновані інтерполяції мають ряд серйозних недоліків: вони не дають можливості характеризувати якість оцінки, налаштування параметрів часто вже не передбачається або робиться приховано від користувача, багато методів нехтують просторовою кореляцією і т. д. Проте розглянемо найбільш поширені детерміновані підходи для просторової інтерполяції.

При використанні детермінованих методів передбачається, що аналізовані дані описуються деякою детермінованою функцією $Z(x, \lambda)$, визначеною на досліджуваній

області S , де $x \in S$ – координати точки; λ - набір внутрішніх параметрів моделі. Завдання полягає в тому, щоб, базуючись на відомих даних ($Z_i = Z(x_i)$ - значення, виміряні в точках $x_i \in S$) і на іншій контекстній інформації про досліджуване явище, підібрати набір параметрів λ і побудувати функцію $Z(x, \lambda)$ для всієї досліджуваної області S . Після цього значення в будь-якій точці просто обчислюється за формулою.

Детерміновані інтерполятори можуть бути глобальними (всі точки з відомими значеннями використовуються при інтерполяції) або локальними (тільки частина значень в точках, найближчих до оцінюваної, використовуються для інтерполяції). Глобальні інтерполятори роблять шукану функцію більш згладженою. При використанні локальних методів околу, яке використовується для оцінки, може здаватися різними способами. Може бути фіксоване число найближчих до оцінюваної точки сусідів, що використовуються при інтерполяції: $N(x) = N = \text{const}$. Тоді розмір зони впливу на значення в точці x , залежить від локальної щільності точок вимірювання. Можливе, навпаки, встановлення розміру області (наприклад, область пошуку D), точки з якої використовуються при оцінюванні значення в x . В цьому випадку $N(x)$ буде змінюватися і в області з рідкісними вимірами при малому значенні D можуть виникнути неоціненні зони (в D -околу точки x недостатньо вимірювань).

В цьому розділі всі методи проілюстровані на даних по радіоактивному забрудненні ґрунту ^{137}Cs .

Лінійні інтерполятори

Лінійні інтерполятори представляють шукану функцію у вигляді лінійної комбінації відомих значень:

$$Z^*(x_0) = \sum_{i=1}^{N(x_0)} w_i(x_0) Z(x_i), \quad (4.9)$$

де $Z^*(x_0)$ - оцінюване значення в точці x_0 ; $Z(x_i)$ - відомі значення в точках вимірювань x_i ; $N(x_0)$ - кількість вихідних точок, які приймають участь в оцінці для координати x_0 ; $w_i(x_0)$ - вагові коефіцієнти. В даному випадку набір параметрів

складається з вагових коефіцієнтів і кількості точок для оцінки. Ці параметри визначаються окремо для кожної точки, що підлягає оцінці.

Лінійні інтерполятори розрізняються формою вагових коефіцієнтів, які задають різні особливості функції. Наприклад, у формі лінійного інтерполятора можна задати полігонний метод Тіссена (метод найближчого сусіда). В цьому випадку вагові коефіцієнти задаються формулою

$$w_i(x) = \begin{cases} 1, & x \in A(x_i), \\ 0, & x \notin A(x_i), \end{cases}$$

де $A(x_i)$ - область впливу точки x_i .

Широко використовується лінійний інтерполятор з ваговими коефіцієнтами, обернено пропорційними відстані до оцінюваної точки в ступені. Вагові коефіцієнти, визначаються за формулою

$$w_i(x) = \frac{\frac{1}{h_i^\beta}}{\sum_{j=1}^n \frac{1}{h_j^\beta}},$$

де β – ступінь; $h_i = \sqrt{d_i^2 + \delta^2}$; d_i – відстань між точками x_i і x_0 ; δ – згладжуючий параметр. Нижня частина дробу у ваговому коефіцієнті вводиться для виконання умови незсуненості оцінки $E\{Z^*(x_0) - Z(x_0)\} = 0$, яке відповідає умові на вагові коефіцієнти $\sum_{i=1}^n w_i = 1$.

Цей метод може бути точним (точне відтворення значень у вихідних точках) і плавним, що характеризується згладжуючим параметром. Точним метод буде при $\delta = 0$. У цьому випадку виникає артефакт у вигляді «бичачих очей» (вигинання до точних значень) навколо точок вимірювань. Згладжує параметр сприяє видаленню цього артефакту.

Як значення ступеня найчастіше використовується значення 2. Такий варіант методу відомий як метод обернених квадратів [276].

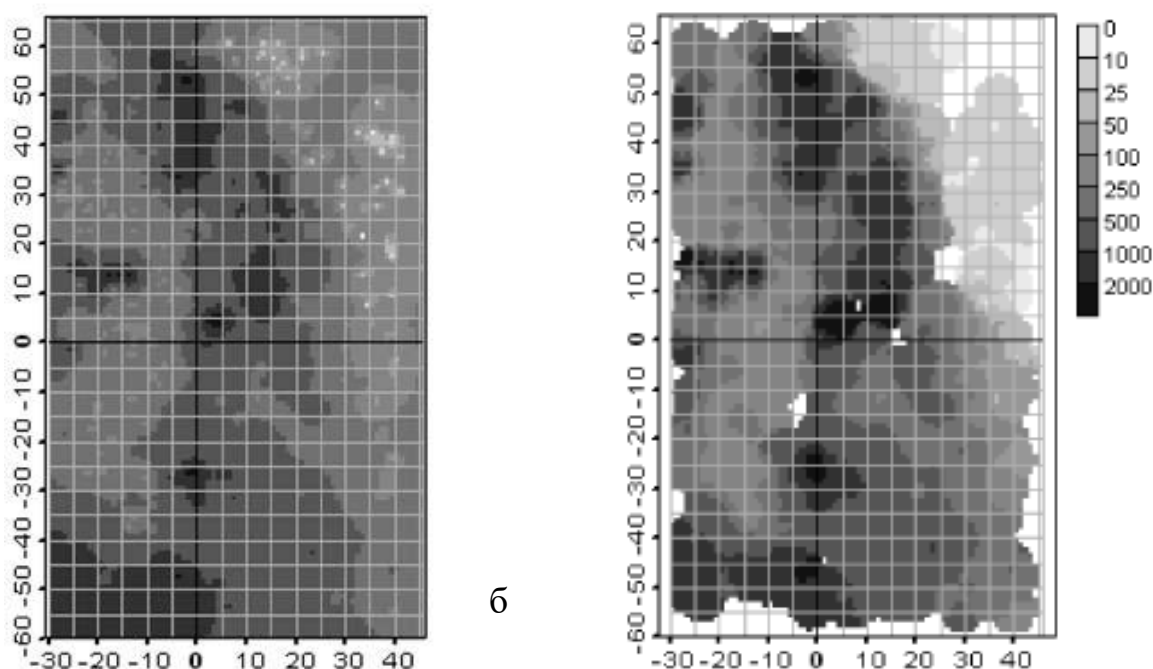


Рис. 4.7. Результат інтерполяції методом обернених квадратів: а - радіус пошуку 50, б - радіус пошуку 5

Проілюструємо метод на даних по радіоактивному забрудненні ґрунту ^{137}Cs .

Приклади використання методу обернених квадратів в глобальному і локальному варіантах наведені на рис. 4.7 для різних значень параметра моделі - радіуса області пошуку $N(x_0)$ сусідніх даних для оцінки (4.6). Чим більший радіус пошуку, тим більш розмита інтерполяційна оцінка (див. Рис. 4.6а). При малому радіусі пошуку оцінка стає більш контрастною, при цьому високі значення не згладжуються (див. Рис. 4.6б).

Проблема сильної залежності найпростішої інтерполяційної оцінки методу обернених квадратів від єдиного параметра - розміру області пошуку - є насправді глибшою. Лінійна регресійна оцінка передбачає наявність певної залежності між даними, які беруть участь в інтерполяції. Це може бути обернена пропорційність квадрату відстані, як в описаному вище методі, або більш складна залежність. Залежності між даними в просторі можуть поширюватися на обмежену відстань. Так, при використанні всіх даних для оцінки в будь-якій точці передбачається, що дані на будь-яких відстанях мають вплив на значення оцінки. У протилежному граничному випадку залежності між даними не існує навіть на мінімальній відстані

між точками. Це означає, що дані розподілені випадково і жоден з методів детермінованої інтерполяції не має сенсу, оскільки при їх використанні передбачається та чи інша залежність.

Таким чином, вибір відповідного радіусу пошуку тісно пов'язаний з поняттями просторової безперервності і просторової кореляції.

Для підбору оптимального значення радіусу пошуку можна використовувати один з алгоритмів перевірки якості оцінки - крос-валідацію або метод складного ножа. При використанні крос-валідації оптимальний радіус пошуку визначається шляхом мінімізації кросс-валідаційної помилки. На цьому принципі засновані деякі алгоритми просторового *автокартографування* [335].

Поліноміальні методи

Поліноміальні інтерполятори [309] представляють значення в точці у вигляді полінома від координат. У двовимірному випадку - для точки x з координатами (x, y) $Z^*(x, y) = P_n(x, y)$, де P_n - поліном n -го ступеня [137].

Зазвичай на практиці для двовимірного випадку використовують один з чотирьох типів поліномів:

- площину: $P_1(x, y) = a + bx + cy$;
- білінійно-сідловий: $P_{1.5}(x, y) = a + bx + cy + dxy; \frac{n!}{r!(n-r)!}$;
- квадратичний: $P_2(x, y) = a + bx + cy + dxy + ex^2 + fy^2$;
- кубічний: $P_3(x, y) = a + bx + cy + dxy + ex^2 + fy^2 + gx^2y + hxy^2 + ix^3 + jy^3$.

Теоретично можна використовувати і поліноми більш високого порядку. Вони визначаються максимальним ступенем при x , максимальним ступенем при y і спільним максимальним ступенем. Всі проміжні ступені в поліномі будуть присутні [137].

Завдання поліноміальної інтерполяції зводиться до того, щоб визначити невідомі коефіцієнти a_i так, щоб поліноми максимально добре відповідали даним в заданих точках. Для цього знаходять мінімум по всім коефіцієнтам (a, b, c, d і т. д.) функції χ^2 , яка задає інтегральну помилку інтерполяції і визначається наступним чином:

$$\chi^2 = \sum_{i=1}^N [Z(x_i, y_i) - P_n(x_i, y_i)]^2.$$

Мінімізація полягає у вирішенні системи лінійних рівнянь з числом невідомих, що дорівнює кількості рівнянь. Число рівнянь (невідомих) залежить від обраного полінома.

Будь-який глобальний поліноміальний метод, взагалі кажучи, не є інтерполятором в суворому сенсі, скоріше він відноситься до апроксиматорів. Його можна використовувати, наприклад, для виділення великомасштабного тренду.

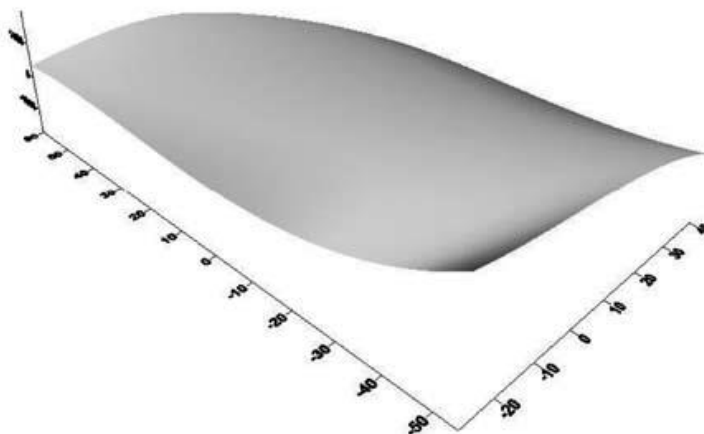
Можна скористатися і локальним варіантом поліноміального методу, коли пошук коефіцієнтів проводиться тільки на основі даних, які потрапили в зону пошуку. Приклади застосування глобального полігону третього порядку і локального варіанту з полігоном другого порядку наведені на рис. 4.7.

Метод базисних функцій

Оцінка методом базисних функцій будується як лінійна комбінація з базисних функцій:

$$Z^*(x_0) = \sum_{i=1}^n c_i B(h_{0i}),$$

де h_{i0} - відстань від точки x_0 до точки x_i ; $B(h_{i0})$ - базисна функція, що визначається від відстані; c_i - вагові коефіцієнти. Коефіцієнт c_i визначає алгебраїчний знак входження відповідного члена та ступінь його впливу. Класичний варіант методу є точним, але можливе введення згладжуючого параметра δ .



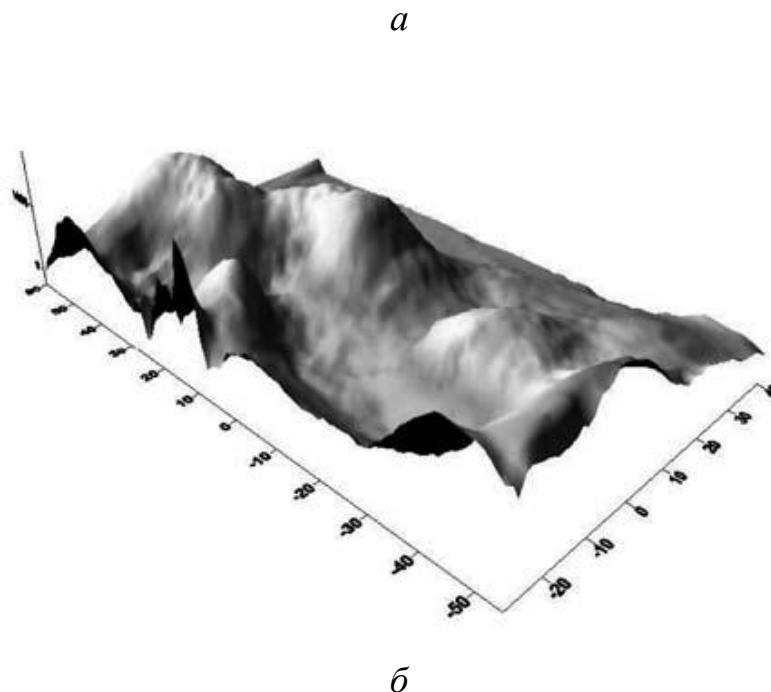


Рис. 4.8. Результат глобальної інтерполяції поліномом третього ступеня (а) і локальної інтерполяції поліномом другого ступеня (б)

Традиційно використовуються наступні типи базисних ядерних функцій:

- обернений мультиквадрік: $B(h) = \frac{1}{\sqrt{h^2 + \delta^2}}$;
- мультилогарифмічний: $B(h) = \lg(h^2 + \delta^2)$;
- мультиквадратичний: $B(h) = \sqrt{h^2 + \delta^2}$;
- природний кубічний сплайн: $B(h) = (h^2 + \delta^2)^{\frac{1}{2}}$;
- тонкий сплайн: $B(h) = (h^2 + \delta^2) \lg(h^2 + \delta^2)$.

Вагові коефіцієнти c_i визначаються з умови точності оцінки в відомих точках, тобто у всіх заданих точках (x_i, y_i) модель інтерполяції повинна давати оцінку, рівну заданим значенням $V(x_i, y_i)$. Визначення ваг проводиться при $\delta = 0$, при використанні згладженого варіанта δ використовується при оцінці. Таким чином, щоб знайти коефіцієнти c_i , необхідно вирішити систему з N лінійних рівнянь з N невідомими:

$$\begin{cases} \sum_{i=1}^N c_i [q(x_i, y_i, x_1, y_1)] = V(x_1, y_1), \\ \vdots \\ \sum_{i=1}^N c_i [q(x_i, y_i, x_N, y_N)] = V(x_N, y_N). \end{cases}$$

Метод базисних функцій володіє універсальністю та ефективністю. Приклад застосування методу базисних функцій (мультіквадратичні ядра) наведено на рис. 4.9.

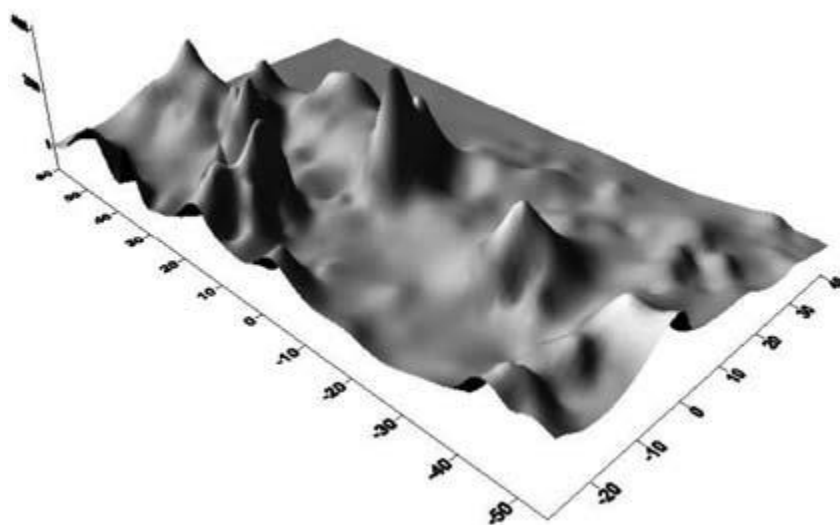


Рис. 4.9. Результат інтерполяції методом базисних функцій (мультіквадратичні ядра)

4.4. Геостатистичні інтерполяції для скалярної функції методом крігінгу

Крігінг - базова інтерполяційна модель геостатистики. Він є основою всіх методів, пов'язаних з геостатистикою, - інтерполяції, імовірнісного картування, стохастичного моделювання. Термін «крігінг» служить для позначення сімейства алгоритмів лінійної просторової регресії. Він походить від прізвища інженера Д. Кріга, який першим застосував інтерполятор на основі моделі просторової кореляції даних для аналізу золотих родовищ Південної Африки [295]. Л. С. Гандіні

незалежно від Д. Кріга застосовував аналогічний метод для об'єктивного аналізу метеополей [25].

Виділяють кілька варіантів моделей крігінга (простий, звичайний, універсальний, логнормальний, невязок і ін.), які розрізняються прийнятими припущеннями і використовуваною інформацією про модельовану зміну.

Основні постулати крігінга

Розглянемо проблему оцінювання значення неперервної змінної Z в довільній точці x , що належить області простору S . Вихідна інформація про змінну представлена у вигляді набору $\{z(x_i), i = 1, \dots, n\}$ з n вимірювань, зроблених в точках x_1, x_2, \dots, x_n простору.

Всі інтерполятори сімейства крігінга є різного роду модифікаціями базового лінійного регресійного оцінювача $Z^*(x)$, що визначається таким чином:

$$Z^*(x) - m(x) = \sum_{i=1}^{n(x)} \lambda_i(x) [Z(x_i) - m(x_i)], \quad (4.10)$$

де $\lambda_i(x)$ - вагові коефіцієнти, що стосуються даних $z(x_i)$. Значення $z(x_i)$ інтерпретуються як реалізації випадкової змінної $Z(x_i)$. Величини $m(x)$ і $m(x_i)$ є математичними очікуваннями випадкових змінних $Z(x)$ і $Z(x_i)$. Число даних, що використовуються при оцінці, і значення вагових коефіцієнтів можуть змінюватися в залежності від місця розташування оцінюваної точки x .

Тип оцінювача залежить від моделі випадкової функції $Z(x)$. Розкладемо її на дві компоненти - детермінований тренд $m(x)$ і випадкову невязку $R(x)$:

$$Z(x) = m(x) + R(x). \quad (4.11)$$

Компонента невязки $R(x)$ моделюється як стаціонарна випадкова функція з нульовим математичним очікуванням $m_R(x)$ і коваріацією $C_R(h)$:

$$E\{R(x)\} = 0,$$

$$\text{Cov}\{R(x), R(x+h)\} = E\{R(x) R(x+h)\} = C_R(h).$$

Математичне сподівання просторової змінної Z в точці x , таким чином, буде дорівнює значенню тренда:

$$E\{Z(x)\} = m(x).$$

Далі ми розглянемо різновиди крігінга для моделювання однієї змінної, які визначаються припущенням про вид тренда.

Всі методи сімейства крігінга використовують одну і ту ж цільову функцію для мінімізації, а саме варіацію помилки оцінки $\sigma_E^2(x)$ при додатковій умові незсуненості оцінки, іншими словами, варіація

$$\sigma_E^2(x) = \text{Var}\{Z^*(x) - Z(x)\}, \quad (4.12)$$

мінімізується при обмеженні

$$E\{Z^*(x) - Z(x)\} = 0. \quad (4.13)$$

Спочатку всі крігінги розглядалися як глобальні оцінювачі, тобто для оцінки значення в точці x_0 з області S передбачалося використовувати всі наявні вимірювання $\{z(x_i), i = 1, \dots, n\}$. Тоді припущення, наприклад про сталість середнього поширюється на всю (можливо, досить велику) область, що, взагалі кажучи, рідко зустрічається в природі. Для ослаблення такого сильного припущення, на практиці зазвичай використовують локальну оцінку на основі $n(x)$ найближчих до точки оцінювання даних. Можна сказати, що використовувані при оцінці дані вибираються з деякого околу $W(x)$ точки оцінювання x . Розмір і форма цього околу залежать від вихідних даних: пропонується використовувати зону, орієнтовану відповідно до еліпсу кореляції, але можливо і меншого або більшого розміру. Зменшення околу дозволяє отримувати більш варіабельну (менш згладжену) оцінку.

Простий крігінг

Простий крігінг (simple kriging - SK) працює в припущенні про стаціонарність другого порядку випадкової змінної $Z(x)$. Крім того, передбачається, що детермінована компонента $m(x)$ в постійна і відома на всій області дослідження S :

$$m(x) = m, \forall x \in S.$$

Знання середнього значення m дає можливість зробити просте перетворення шляхом вирахування постійного тренда

$$Y(x) = Z(x) - m, \quad (4.14)$$

і далі будувати лінійний оцінювач для випадкової функції $Y(x)$ на всій області S :

$$Y^*(x) = \sum_{i=1}^{n(x)} \lambda_i(x) Y(x_i), \quad (4.15)$$

автоматично отримуючи незміщеність оцінки (збереження глобального середнього). Так як $E\{Y(x)\} = 0, \forall x \in S$, то

$$E\{Z^*(x) - Z(x)\} = E\{Y^*(x) - Y(x)\} = E\left\{\sum_{i=1}^{n(x)} \lambda_i(x) Y(x_i) - Y(x)\right\} = 0.$$

Остаточна оцінка простого крігінга з (4.14) і (4.15) має вигляд

$$Z^*(x) = m + \sum_{i=1}^{n(x)} \lambda_i(x) Y(x_i). \quad (4.16)$$

Тепер розглянемо варіацію помилки σ_R^2 для оцінки функції $Y(x)$.

$$\begin{aligned} \sigma_R^2 &= \text{Var}\{Y^*(x) - Y(x)\} = E\{(Y^*(x) - Y(x))^2\} = \\ &= \text{Var}\{Y^*(x)\} - 2\text{Cov}\{Y^*(x)Y(x)\} + \text{Var}\{Y(x)\}. \end{aligned} \quad (4.17)$$

Так як функція $Z(x)$ задовольняє стаціонарності другого порядку, цій умові задовольняє і функція $Y(x)$. Тоді всі коваріації і варіації, що входять в (4.17), існують. Щоб отримати варіацію оцінки, підставимо в перше і друге доданки суми (4.17) формулу оцінки (4.18):

$$\text{Var}\{Y^*(x)\} = \sum_{i=1}^{n(x)} \sum_{j=1}^{n(x)} \lambda_i(x) \lambda_j(x) C_{ij}, \quad (4.18)$$

$$2\text{Cov}\{Y^*(x)Y(x)\} = 2 \sum_{i=1}^{n(x)} \lambda_i(x) C_{i0}, \quad (4.19)$$

де $C_{ij} = \text{Cov}\{Y(x_i)Y(x_j)\}$, $C_{i0} = \text{Cov}\{Y(x_i)Y(x)\}$.

Варіація невідомої випадкової змінної $Y(x)$ також існує і пов'язана з апіорною варіацією вихідних даних σ_Z^2 :

$$\text{Var}\{Y(x)\} = \text{Var}\{Z(x)\} - m^2 = \sigma_Z^2 - m^2 = \sigma_Y^2. \quad (4.20)$$

В результаті отримаємо значення варіації помилки оцінки змінної Y як суму (4.18), (4.19) і (4.20):

$$\sigma_R^2(x) = \sum_{i=1}^{n(x)} \sum_{j=1}^{n(x)} \lambda_i(x) \lambda_j(x) C_{ij} - 2 \sum_{i=1}^{n(x)} \lambda_i(x) C_{i0} + \sigma_Y^2. \quad (4.21)$$

Крігінг як найкращий оцінювач з класу лінійних повинен мати мінімальну варіацію помилки. Вагові коефіцієнти $\lambda_i(x)$ в (4.21) підбираються так, щоб вони

мінімізували варіацію помилки (4.22), тобто щоб похідна від варіації по всіх вагах дорівнювала нулю. В результаті диференціювання виходить система рівнянь простого крігінга - лінійна система з $n(x)$ рівнянь з $n(x)$ невідомими:

$$\sum_{j=1}^{n(x)} \lambda_j(x) C_{ij} = C_{i0}, \forall i = 1, \dots, n(x). \quad (4.22)$$

Система рівнянь простого крігінга (4.22) має єдине рішення, якщо матриця коваріацій несингулярна. Це умова виконана при позитивній визначеності функції коваріації і відсутності серед набору вихідних точок $x_1, \dots, x_{n(x)}$, які просторово збігаються або дуже близько розташовані. Точки, які збігаються або близькі формують лінійно залежні рядки матриці C_{ij} .

Оцінка функції $Z(x)$ виходить підстановкою отриманих вагових коефіцієнтів в формулу (4.16).

Помилка оцінки простого крігінга (варіація простого крігінга) виходить з формули (4.21) підстановкою в неї (4.22). Варіацію простого крігінга можна обчислити за формулою

$$\sigma_{SK}^2(x) = \sigma_Z^2 - \sum_{i=1}^{n(x)} \lambda_i(x) c_{i0}. \quad (4.23)$$

Простий крігінг має низку властивостей.

- Оцінка простого крігінга є точною. Це означає, що якщо координата оцінюваної точки x_0 збігається з якоюсь координатою з вихідного набору даних ($x_0 = x_i, i = 1, \dots, n$), то отримана оцінка буде також збігатися з вихідним значенням: $Z^*(x_0) = Z(x_k)$. Це легко довести, користуючись одиничністю рішення системи рівнянь простого крігінга.

- Ваги простого крігінга не залежить від значень вихідного набору даних, а залежать тільки від просторової кореляції поля, побудованого на основі даних. Таким чином, якщо є кілька наборів вихідних даних, виміряних в одних і тих же точках і описуваних однаковими (або мультиплікативно пов'язаними) функціями коваріації, то для обчислення оцінки простого крігінга в загальній точці x_0 систему рівнянь простого крігінга досить вирішити один раз, а потім використовувати отримані ваги для всіх змінних.

- Оцінка простого крігінга є згладженої в порівнянні з розподілом вихідних даних. Як видно з (4.23), варіація оцінки простого крігінга σ_{SK}^2 менше значення варіації вихідних даних σ_2 .

- Помилка простого крігінга ортогональна оцінці простого крігінга в гільбертовому просторі, побудованому з усіх можливих лінійних комбінацій вихідних даних і має в якості метрики коваріацію. Ця властивість зайвий раз підтверджує, що простий крігінг є найкращою оцінкою в класі лінійних оцінювачей.

Основним недоліком простого крігінга є припущення про знання середнього. Використання в якості середнього його статистичної оцінки (математичного очікування) робить ваги залежними від значень вихідного набору даних. Крім того, оцінка математичного очікування може виявитися спотвореною, зміщеною і т. д., наприклад при високій кластерності вихідної мережі моніторингу. Тому простий крігінг рідко застосовується як самостійний метод оцінювання, зазвичай його використання пов'язане з штучними комбінаціями, де середнє відомо внаслідок попередніх маніпуляцій з вихідними даними.

Звичайний крігінг

Звичайний крігінг (ordinary kriging - ОК) відрізняється від простого крігінга тим, що не передбачає знання середнього значення. У звичайному крігінгу середнє значення вважається постійним, але воно невідомо. Крім того, звичайний крігінг при використанні локальної оцінки не вимагає сталості середнього по всій зоні оцінювання; передбачається, що середнє постійне тільки в околі точки оцінювання $W(x)$. Припущення про сталість середнього в рамках малого околу більш реалістичне, тим більше що дані володіють просторовою безперервністю.

Оцінка звичайного крігінгу будується, як лінійна комбінація вихідних даних:

$$Z^*(x) = \sum_{i=1}^{n(x)} \lambda_i(x) Z(x_i). \quad (4.24)$$

Розглянемо умову незміщення (4.24) в разі невідомого середнього:

$$E\{Z^*(x) - Z(x)\} = E\left\{\sum_{i=1}^{n(x)} \lambda_i(x) Z(x_i) - Z(x)\right\} = \left[\sum_{i=1}^{n(x)} \lambda_i(x) - 1\right] m = 0,$$

тобто умова незміщення буде виконана, якщо сума ваг, що використовуються при оцінці, дорівнює одиниці:

$$\sum_{i=1}^{n(x)} \lambda_i(x) = 1, \quad (4.25)$$

Таким чином, відсутність знання про значення середнього накладає на ваги $\lambda_i(x)$ додаткову вимогу. Тепер, щоб виконувалась властивість найкращого оцінювача, потрібно знаходити ваги, які мінімізують варіацію при додатковому обмеженні (4.24).

Вирішення цього завдання здійснюється з використанням мінімізації лагранжіана $L(x)$, куди крім варіації включається умова (4.25) з ваговим коефіцієнтом $\mu(x)$ (множником Лагранжа):

$$L(x) = \sum_{i=1}^{n(x)} \sum_{j=1}^{n(x)} \lambda_i(x) \lambda_j(x) C_{ij} - 2 \sum_{i=1}^{n(x)} \lambda_i(x) C_{i0} + \sigma_Z^2 + 2\mu(x) \left(\sum_{i=1}^{n(x)} \lambda_i(x) - 1 \right),$$

де C_{ij} – коваріація випадкових змінних:

$$C_{ij} = \text{Cov}\{Z(x_i)Z(x_j)\}, i, j = 1, \dots, n,$$

$$C_{i0} = \text{Cov}\{Z(x_i)Z(x)\}, \quad i = 1, \dots, n.$$

Для мінімізації лагранжіана $L(x)$ необхідно його продиференціювати по вагам $\lambda_i(x)$ і коефіцієнту $\mu(x)$, а потім прирівняти ці похідні нулю. В результаті отримуюмо лінійну систему з $n(x) + 1$ рівнянь з $n(x) + 1$ невідомими - система рівнянь звичайного крігінга:

$$\begin{cases} \sum_{j=1}^{n(x)} \lambda_j(x) C_{ij} + \mu(x) = C_{i0}, i = 1, \dots, n(x), \\ \sum_{i=1}^{n(x)} \lambda_i(x) = 1. \end{cases} \quad (4.26)$$

Система рівнянь (4.26) аналогічно з системою рівнянь простого крігінга (4.22) має єдине рішення при позитивній визначеності функції коваріації C і відсутності точок, які просторово збігаються або дуже близькі.

Для обчислення оцінки знайдені ваги $\lambda_i(x)$ підставляються в лінійну комбінацію (4.27). Варіація звичайного крігінга обчислюється з формули (4.27) з використанням першої частини системи (4.26):

$$\sigma_{OK}^2(x) = \sigma_Z^2 - \sum_{i=1}^{n(x)} \lambda_i(x) c_{i0} + \mu(x). \quad (4.27)$$

На практиці частіше замість припущення про стаціонарність другого порядку і функції коваріації користуються менш слабким припущенням про внутрішню гіпотезу і пов'язану нею варіограму. Система рівнянь звичайного крігінга (4.27) легко може бути переписана в термінах варіограми:

$$\begin{cases} \sum_{j=1}^{n(x)} \lambda_j(x) \gamma_{ij} + \mu(x) = \gamma_{i0}, i = 1, \dots, n(x), \\ \sum_{i=1}^{n(x)} \lambda_i(x) = 1. \end{cases}$$

Варіація звичайного крігінга (4.28) також може бути переписана в термінах варіограми:

$$\sigma_{OK}^2(x) = \sigma_Z^2 - \sum_{i=1}^{n(x)} \lambda_i(x) \gamma_{i0} + \mu(x). \quad (4.28)$$

Всі властивості, описані для простого крігінга, відносяться в тій же мірі і до звичайного крігінга. Але порівняння формул варіацій простого (4.23) і звичайного (4.27), (4.28) крігінгов показує, що платою за невідоме значення середнього є збільшення варіації, що веде до росту невизначеності оцінки.

Для того щоб зрозуміти, як впливає на оцінку і варіацію крігінга відсутність знання середнього, розглянемо штучний приклад. Нехай відомі одне значення $Z(x_i)$ функції $Z(x)$ і модель просторової кореляції, задана коваріацією $C(h)$ або варіограмою $\gamma(h)$. Використовуючи ці дані, побудуємо оцінку функції в точці x_0 . Якщо відомо середнє (нехай для простоти це буде нуль), модель простого крігінга будується наступним чином:

1. Рівняння простого крігінга:

$\lambda C_{11} = C_{10}$, де C_{11} - значення апіорної варіації; C_{10} – значення коваріаційної функції для вектора, що розділяє точки x_1 і x_0 .

2. Ваговий коефіцієнт крігінга

$$\lambda = \frac{C_{10}}{C_{11}}.$$

3. Визначається оцінка:

$$Z(x_0) = \frac{C_{10}}{C_{11}} Z(x_1).$$

Значення оцінки залежить від взаємної просторової орієнтації точок через просторову кореляцію.

4. Варіація крігінга

$$\sigma^2 = C_{00} - \frac{C_{10}}{C_{11}} C_{10}$$

також визначається просторовою кореляцією.

Універсальний крігінг

Універсальний крігінг, або крігінг з трендом (universal kriging, UK), передбачає, що невідоме середнє значення $m(x)$ плавно змінюється у всій області дослідження S . У деяких випадках неможливо припустити локальну сталість середнього навіть в оточенні оцінюваної точки $W(x)$. Одним з можливих в такому випадку підходів є саме універсальний крігінг. Припускається, що детермінована компонента випадкової змінної (тренд) моделюється як лінійна комбінація $K + 1$ базисних (відомих) функцій $f_k(x)$ (за прийнятою згодою $f_0(x)=1$) з коефіцієнтами $a_k(x)$, невідомими і постійними всередині околу оцінюваної точки $x \in W(x)$:

$$m(x') = \sum_{k=0}^K a_k(x') f_k(x'), \quad a_k(x') = a_k, \quad \forall x' \in W(x). \quad (4.29)$$

Розглянемо, як виконати в такому випадку умову незсуненості оцінки:

$$\begin{aligned} E\{Z_{UK}^*(x) - Z(x)\} &= \sum_{i=1}^{n(x)} \lambda_i(x) m(x_i) - m(x) = \\ &= \sum_{i=1}^{n(x)} \lambda_i(x) \sum_{k=0}^K a_k(x) f_k(x_i) - \sum_{k=0}^K a_k(x) f_k(x) = \\ &= \sum_{k=0}^K a_k(x) \left[\sum_{i=1}^{n(x)} \lambda_i(x) f_k(x_i) - f_k(x) \right]. \end{aligned}$$

Отримуємо набір з $K + 1$ додаткових обмежень, але позбавляємося від необхідності оцінювати коефіцієнти $a_k(x)$:

$$\sum_{i=1}^{n(x)} \lambda_i(x) = 1,$$

$$\sum_{i=1}^{n(x)} \lambda_i(x) f_k(x_i) = f_k(x), k = 1, \dots, K.$$

Побудувавши відповідний лагранжіан, продиференціювавши його по всім невідомим змінним і прирівнявши до нуля відповідні похідні, отримуємо систему рівнянь універсального крігінга:

$$\begin{cases} \sum_{j=1}^{n(x)} \lambda_j(x) C_R(x_i - x_j) + \sum_{k=0}^K \mu_k(x) f_k(x_i) = C_R(x_i - x), & i = 1, \dots, n(x) \\ \sum_{j=1}^{n(x)} \lambda_j(x) = 1, \\ \sum_{j=1}^{n(x)} \lambda_j(x) f_k(x_j) = f_k(x), & k = 1, \dots, K. \end{cases} \quad (4.30)$$

Варіація універсального крігінга також може бути записана наступним чином:

$$\sigma_{UK}^2(x) = C_R(0) - \sum_{i=1}^{n(x)} \lambda_i(x) C_R(x_i - x) - \sum_{k=0}^K \mu_k(x) f_k(x).$$

Тут слід звернути увагу, що в системі рівнянь універсального крігінга використовується коваріації $C_R(\cdot)$ для випадкової компоненти $R(x)$ функції $Z(x)$, апріорне знання яких передбачається. Крім того, потрібне апріорне знання набору базисних функцій. Детальніше ознайомитися з тим, як на практиці підходять до вирішення завдань підготовки моделі тренда і коваріаційної функції залишків, можна, наприклад, в [Goovaerts, 1997; Armstrong, 1984]. Але, взагалі кажучи, універсальний крігінг не отримав широкого поширення, так як завдання підбору функцій для моделювання тренда не є прозорим.

Логнормальний крігінг

Логнормальним випадковим процесом $\{Z(x): x \in S\}$ називається такий позитивно-визначений процес, коли $Y(x)$ є гаусовим процесом:

$$Y(x) \equiv \log Z(x) \sim N(m(x), \sigma^2), \quad x \in S.$$

Нехай $Y^*(x_0)$ - оцінка функції $Y(x)$ в точці x_0 , де немає вимірювання, отримана за допомогою крігінга на підставі відомих даних $Y(x_1), Y(x_2), \dots, Y(x_n)$ в точках вимірювань x_1, x_2, \dots, x_n . Тепер потрібно отримати оцінку функції $Z(x)$ в цій точці. Якщо отримувати таку оцінку, просто роблячи зворотне логарифму перетворення

$$Z^*(x_0) = \exp\{Y^*(x_0)\},$$

то вона буде зміщеною [Dowd, 1982], тобто

$$E\{Z^*(x_0)\} \neq E\{Z(x_0)\}.$$

А мета будь-якого крігінга, в тому числі і логнормального, - найкраща незміщена оцінка. Тому, щоб врахувати і виправити зміщення, що виникають при зворотньому перетворенні, використовується формула

$$Z^*(x_0) = \exp\left\{Y^*(x_0) + \frac{1}{2} [Var\{Y(x_0)\} - Var\{Y^*(x_0)\}]\right\}.$$

Тут варіація $Var\{Y^*(x_0)\}$ є не варіацією звичайного крігінга $\sigma_{OK}^2(x_0)$, а значенням, яке визначається як

$$Var\{Y^*(x_0)\} = \lambda^T \Sigma \lambda,$$

де Σ - варіаційно-коваріаційна матриця значень $Y = [Y(x_1), Y(x_2), \dots, Y(x_n)]^T$. Таким чином, не зміщена оцінка $Z(x_0)$ може бути отримана на підставі наступної формули:

$$\begin{aligned} Z^*(x_0) &= \exp\left\{Y^*(x_0) + \frac{1}{2} [Var\{Y(x_0)\} - Var\{Y^*(x_0)\}]\right\} = \\ &= \exp\left\{Y^*(x_0) + \frac{1}{2} \sigma_{OK}^2(x_0) - \mu\right\}, \end{aligned} \quad (4.30)$$

де μ - множник Лагранжа, значення якого знаходиться при вирішенні системи рівнянь звичайного крігінга для оцінки $Y^*(x_0)$. На практиці логнормальний крігінг зазвичай використовується для даних, де значення розподіляються на порядки. Таке сильне розходження не дає можливості отримати модель просторової

кореляції. Нелінійне логарифмічне перетворення робить дані придатними для геостатистики.

Блоковий і точковий крігінг

Будь-яке вимірювання $Z(x_\alpha)$ завжди співвідноситься з деяким ненульовим кінцевим об'ємом. Це може бути шматок породи або ґрунту, де береться проба. Зазвичай розмір вимірювання дозволяє приписати його до точки з координатою x_α . Розмір оцінюваного значення такий же, як і вимірюваного, і оцінка також приписується до координати. Але іноді метою оцінки є середнє за певним обсягом, якщо саме такий розмір відповідає певним діям (наприклад, очищення і т. д.). *Блоковий крігінг* - узагальнена назва методу для визначення середнього значення функції z по якомусь вимірному сегменту (довжині, площі, обсягу) будь-якого розміру або форми на противагу *точкового крігінгу*, який відноситься до нульового розміру оцінки (проби).

Блокова оцінка може виконуватися, наприклад, як усереднення точкових оцінок крігінга, що потрапили в необхідний обсяг. Інший підхід полягає в використанні безпосередньо системи рівнянь блочного крігінга. Система рівнянь блочного крігінга виглядає точно так, як система рівнянь звичайного крігінга (5.17). Єдина відмінність полягає в тому, що в правій частині системи стоїть коваріація не для двох точок, а для блоку (обсягу) і точки $\bar{C}(x_\alpha, V(x))$. Визначається така коваріація відповідно до формули [Journel, Huijbregts, 1978]

$$\bar{C}(x_\alpha, V(x)) = Cov\{Z(x_\alpha), Z_V(x)\} = \frac{1}{|V|} \int_{V(x)}^N C(x_\alpha - x') dx'.$$

На практиці така коваріація визначається як середнє точкових коваріацій точки x_α і N точок x_i , які дискретно описують обсяг $V(x)$:

$$\bar{C}(x_\alpha, V(x)) = \frac{1}{N} \int_{i=1}^N C(x_\alpha - x_i).$$

Коваріації можуть бути узагальнені на випадок, коли самі вимірювання теж представлені об'ємом. Це істотно, якщо розмір вимірювання можна порівняти з розміром області дослідження і точковий крігінг є некоректним.

4.5. Застосування трендових геопросторових статистичних моделей для оцінки сталого регіонального розвитку України

Для проведення дослідження було використано набір даних, який описує сталий розвиток України в 2013 році Світового центру даних з геоінформатики та сталого розвитку [65]. З метою перевірки гіпотези про існування просторового тренду в розподілі даних про сталий розвиток було використано спочатку картографічний метод, який дозволяє оцінити загальний паттерн просторового розподілу значень. З цією метою було побудовано ряд тематичних карт значень індексу сталого розвитку та його складових за областями на основі поділу на п'ять груп за природними межами класифікації значень. Аналіз карт дозволяє стверджувати про наявність певної зміни у розподілі значення індексу сталого розвитку у географічному просторі. Для підтвердження цього припущення було використано метод побудови поліноміальної трендової поверхні, яка була побудована на основі точкового файлу, який містить дані про геометричні центри областей України. В результаті проведення експериментальних побудов поверхонь було з'ясовано, що практичне значення при виявленні трендів у даних мають поверхні утворені поліномами 1 – 3 ступенів. Якщо поліноми першого ступеня допомагають встановити загальні характеристики тренду даних, то поліноми 2 – 3 ступенів дозволяють підтвердити існування просторової тенденції у даних та визначити загальні риси її розподілу (Додаток А).

Аналіз поліноміальної поверхні 1, 2 порядків для індексу сталого розвитку дозволяє впевнено стверджувати про існування географічного тренду в розподілі індексу. Значення індексу мають загальну тенденцію до зменшення з Північного Заходу на Південний Схід. Звичайно, в цьому розподілі даних існують відхилення, які пов'язані, наприклад, з високим значенням індексу сталого розвитку для м. Київ, але вони тільки підкреслюють наявність певного географічного тренду (Рис.4.10).

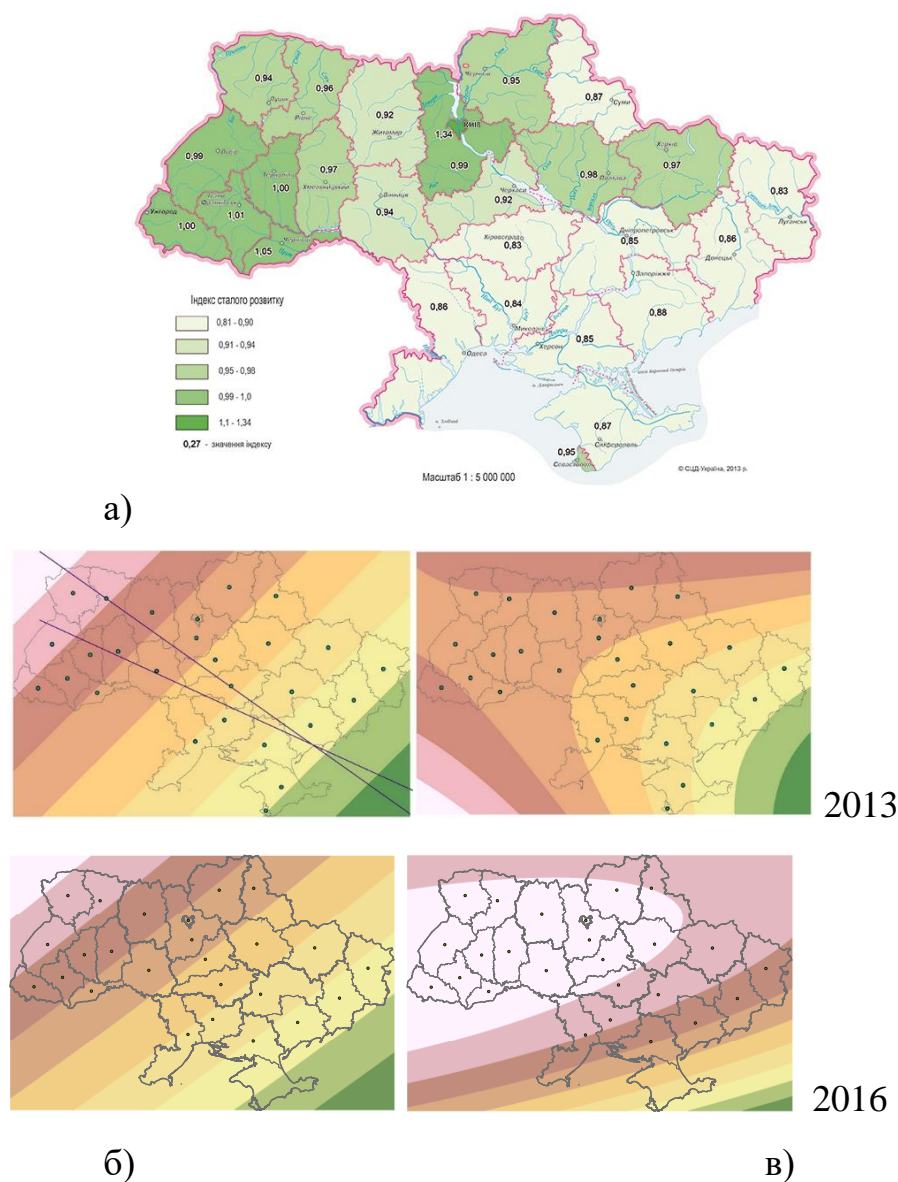


Рис.4.10. Розподіл трендових моделей для індексу сталого розвитку: а – картограма за регіонами; б – поліном 1 порядку в 2013, 2016 рр.; в – поліном 2 порядку в 2013, 2016 рр.

Допоміжним інструментом аналізу тренду є графік аналізу тренду, який у площинах північ – південь, захід – схід демонструє наявність трендів у розподілі даних, які у вигляді результуючого тренду визначаються на трендовій поверхні (Рис.4.30). Таким чином, можна зробити припущення, що західні області України мають тенденцію до більш сталого розвитку, ніж області сходу [137].

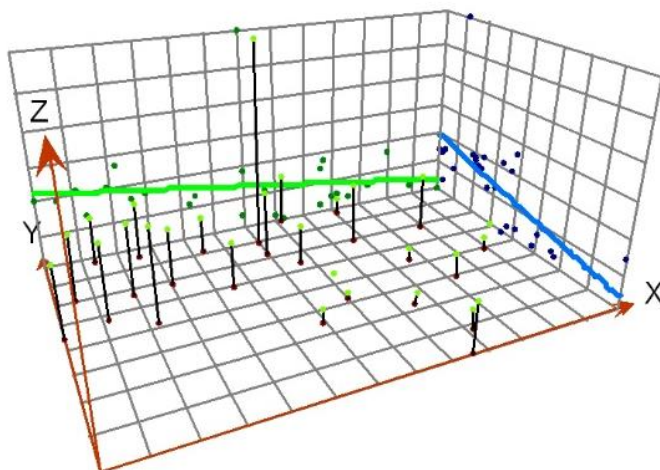


Рис.4.11. Графік аналізу трендової поверхні індексу сталого розвитку

Трендові поверхні можуть бути оцінені як за допомогою значень коефіцієнтів, які використовувались у поліномі, так і описом фізичних характеристик трендової поверхні. Такими характеристиками є ухил поверхні, кут між лініями скату поверхонь першого та другого поліномів, експозиція схилів. Кути нахилу прямо пропорційно та кути між лініями скату обернено пропорційно вказують на наявність вираженого тренду в даних. Кореляційний аналіз між цими показниками вказує на існування тісного зв'язку, який за індексом кореляції Пірсона дорівнює 0,816. Додатковими інструментами аналізу можуть виступати локальні методи інтерполяції даних, такі як обернено-зважені відстані та розподіл ухилів отриманих статистичних поверхонь [137].

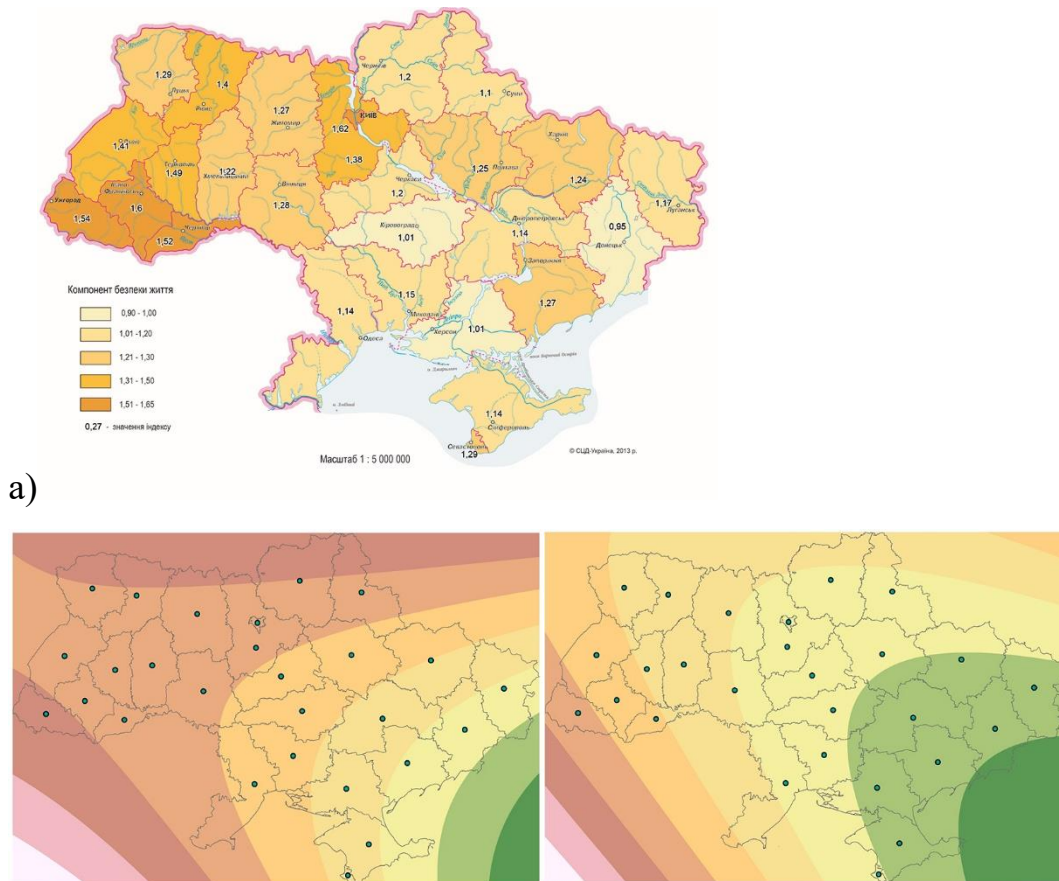
Таблиця 4.1.

Картометричні показники трендових поверхонь індексів

Індекс	Кут нахилу	Кут між лініями скату
<i>Is</i>	0,000004	15,3142
<i>Ie</i>	0,000005	7,6633
<i>Cql</i>	0,000009	23,3202
<i>Isd</i>	0,000011	9,1149
<i>Csl</i>	0,000014	27,1387
<i>Iec</i>	0,000018	54,5763

Інтерпретація виявлених просторових трендів завжди пов'язана зі складностями суб'єктивного та об'єктивного характеру, які викликані пошуком причинно-наслідкових зв'язків. У випадку аналізу просторового тренду у розподілі індексу сталого розвитку можна виходити із ієрархічної структури формування індексу, де підсумкове значення залежить від безпеки і якості життя, які в свою чергу складаються з екологічного, економічного та соціального вимірів та сукупності загроз, які впливають на безпеку життя. З цією метою було проведено відповідний аналіз на виявлення трендів для індексу якості та безпеки життя та трьох вимірів, які є складовими якості життя [137].

Індекс безпеки життя розраховується на підставі даних про 12 ключових загроз, що несуть суспільну небезпеку. Побудовані за його даними поліноміальні поверхні першого та другого порядку вказують на чітку наявність тренду, який дуже тісно збігається з трендом індексу сталого розвитку. Тренд характеризується чітким напрямом з заходу на схід із трохи більшим кутом між лініями скату, ніж у випадку індексу сталого розвитку (Рис.4.12) [137].



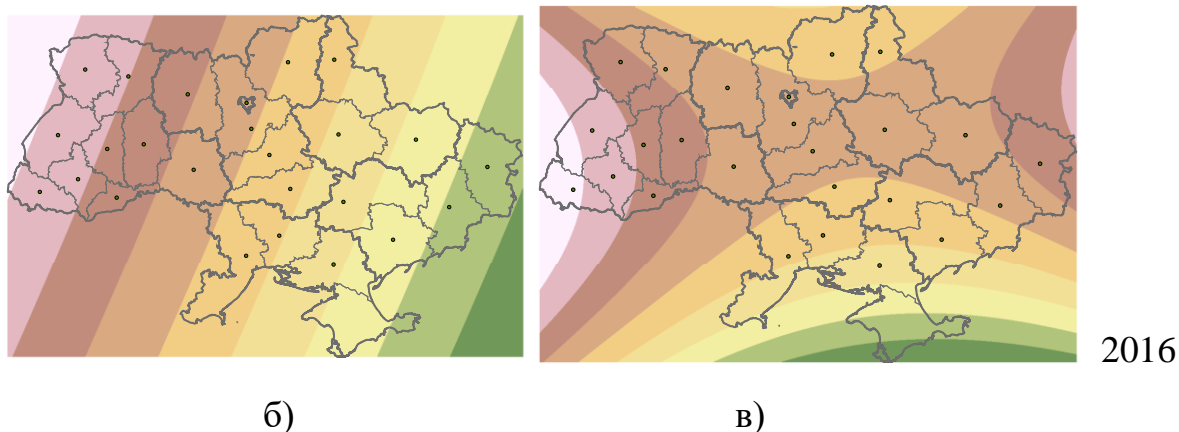
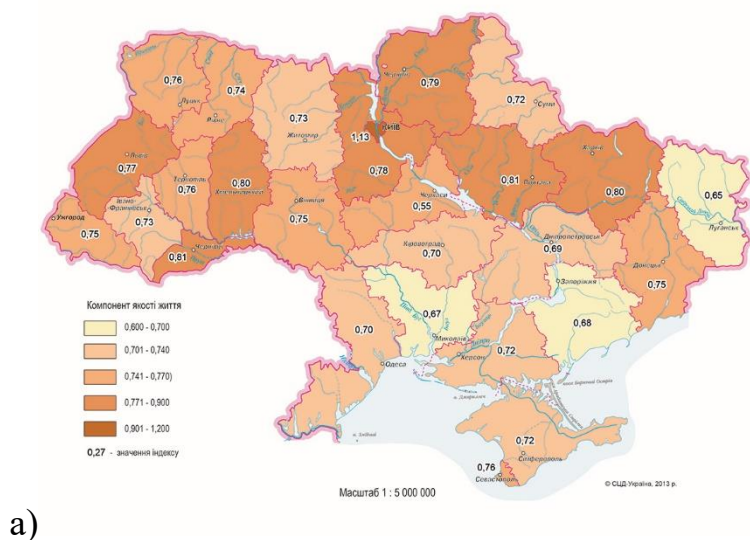


Рис.4.12. Розподіл трендових моделей для індексу безпеки життя: а – картограма за регіонами; поліном 1 порядку в 2013, 2016 рр.; в – поліном 2 порядку в 2013, 2016 рр.

Індекс якості життя, який утворено на основі економічного, соціального та екологічного вимірів характеризується трендом з півночі на південь країни для поліному 1 ступеня та трендом з північного заходу на південний схід для поліному 2 ступеня (Рис.4.13). При цьому трендова поверхня 2 ступеня характеризується значним плато у центральній частині України, де вплив тренду у даних нівелюється іншими факторами. В результаті аналізу двох складових сталого розвитку можна зробити висновки, що для даних індексів якості та безпеки життя тренди також простежуються. При чому індекс безпеки життя має більш виражений тренд даних, ніж індекс якості життя [137].



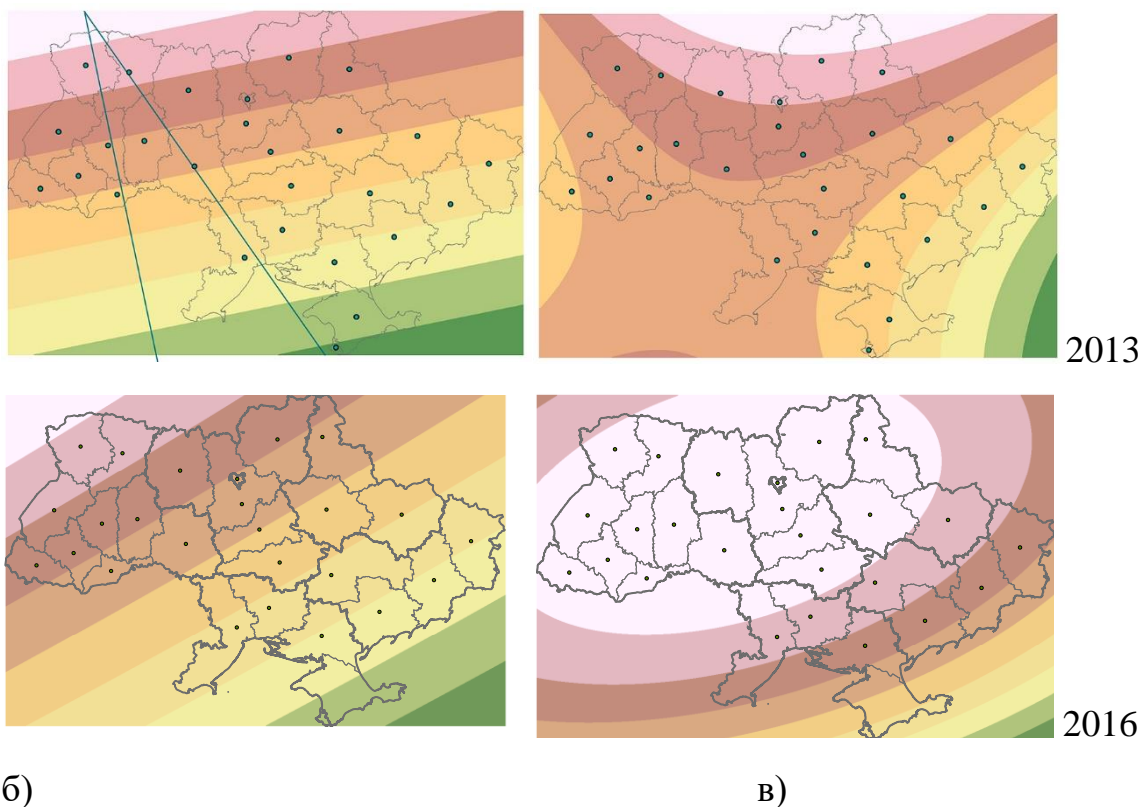


Рис.4.13 Розподіл трендових моделей для індексу якості життя: поліном 1 порядку в 2013, 2016 рр.; в – поліном 2 порядку в 2013, 2016 рр.

Аналіз складових якості життя дозволяє виявити ту групу показників, яка найбільше впливає на формування тренду в даних. Екологічний вимір якості життя має виражений тренд із заходу на схід країни, який візуально видно на трендових поверхнях першого та другого ступенів (Рис.4.33а). Це закономірно свідчить про те, що на заході країни більш сприятлива екологічна ситуація, яка пов'язана з відсутністю у більшості випадків великих джерел антропогенного забруднення та кращим забезпеченням водними ресурсами [137].

Трендові поверхні індексу економічного виміру демонструють зміну з південного заходу на північний схід, що пов'язано зі збільшенням економічних показників на сході країни, де сконцентровані більші виробничі потужності. Проте у випадку аналізу трендових поверхонь першого та другого ступенів існують певні розбіжності і зона плато без змін над територією центральної та східної України. Значною мірою це викликано впливом столичного статусу Києва, який дає найбільший внесок в економічний розвиток (Рис. 4.33б).

Аналіз третього соціального виміру в структурі індексу якості життя населення показав відсутність єдиного тренду в даних соціальної складової для території України. Це демонструє досить строкату структуру соціального розвитку України, яка має плато значень у центральній частині України та два тренди до зменшення показників на північному заході та південному сході країни (Рис. 4.14в). Таким чином, гіпотеза про наявність вираженого географічного тренду в даних про соціальний розвиток не пройшла перевірку. В даних існують тільки окремі тенденції, які можуть бути визначені за допомогою локальних методів інтерполяції [137].

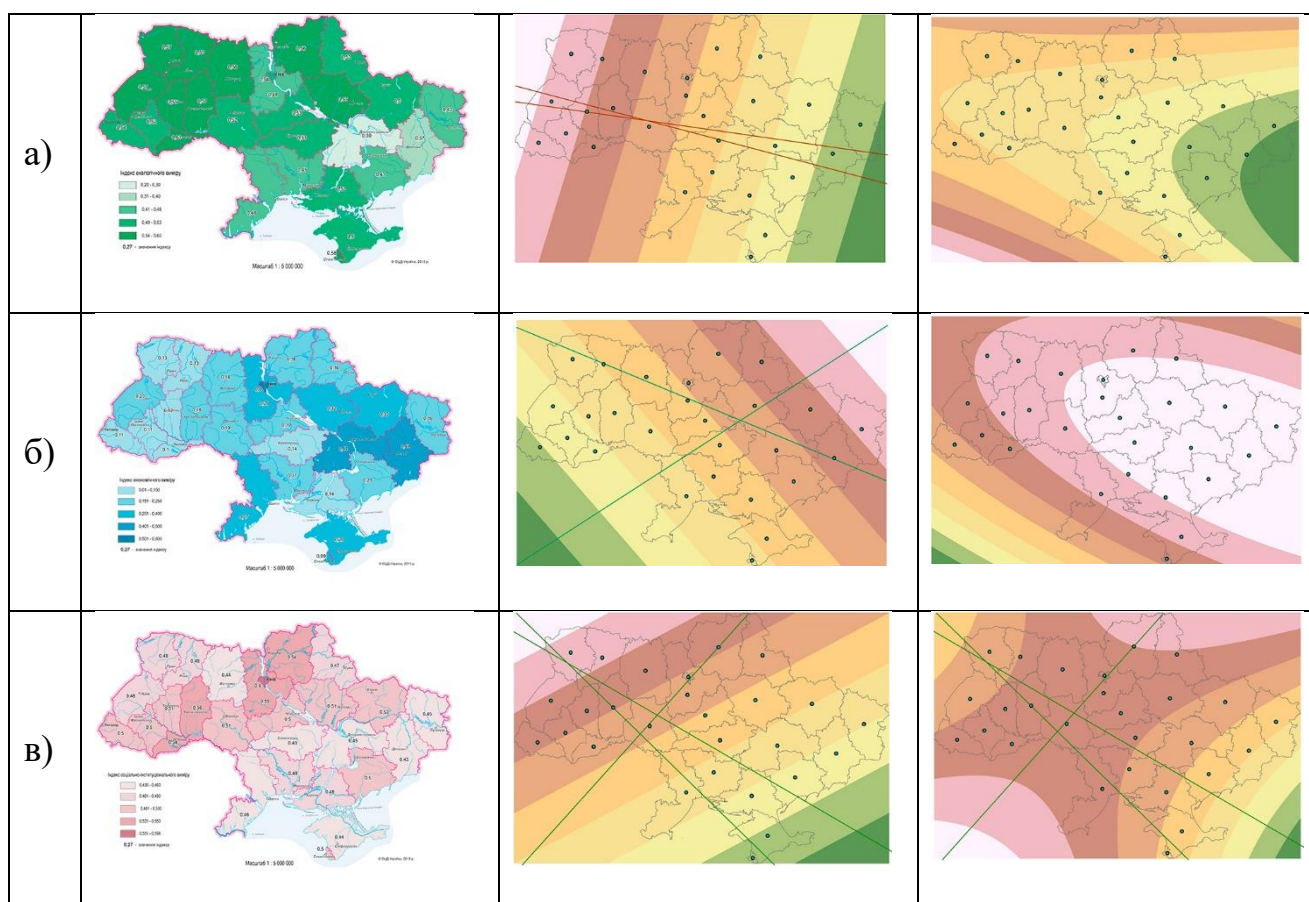


Рис. 4.14. Розподіл трендових моделей для: а – екологічного виміру, б – економічного виміру, в – соціально-інституціонального виміру.

В результаті можна зазначити, що аналіз складових індексу якості життя населення показує достатньо різноманітну структуру глобальних та локальних географічних трендів в структурі даних. Найбільш виражений тренд наявний в даних екологічного виміру, що пов'язано з особливостями природокористування в

різних регіонах країни. Більш складну структуру має трендова поверхня другого ступеня для економічного виміру, яка детермінована регіональними диспропорціями в економічному розвитку та промисловому виробництві. При цьому дані екологічного та економічного вимірів мають протилежні за напрямом змін тренди. Перевірка соціального виміру на наявність глобальних трендів показала, що для соціального виміру в Україні глобальні просторові тренди відсутні і є тільки локальні регіональні тенденції, які можуть бути описані за допомогою локальних методів інтерполяції [137].

Методи детекції геопросторових трендів в ієрархічних моделях даних

На основі картографічних методів аналізу встановлено ступінь впливу складових індексу сталого розвитку на формування тренду інтегрального показника. Таблиця 2 демонструє сукупний аналіз складових індексу сталого розвитку з використанням трьох методів: картографічний, трендова поверхня 1 ступеня, трендова поверхня 2 ступеня. Уся сукупність станів розподілу даних зведена до розмірності $f(x) \in [0;1]$, де $f(x)$ - функція наявності тренду даних. В цьому діапазоні $y = 0 \Rightarrow f(x) = 0$, $y = 0,5 \Rightarrow f(x) \approx 1$, $y = 1 \Rightarrow f(x) = 1$.

Це означає, що значення 0 вказує на відсутність глобального географічного тренду, 0,5 – можливість наявності анізотропії у розподілі географічних даних, 1 вказує на виражений географічний тренд у даних.

Таблиця 4.2.

Визначення географічного тренду у розподілі даних [137]

Показник	Картографічний метод	Поліном 1 ступеня	Поліном 2 ступеня
Індекс сталого розвитку	1	1	1
Якість життя	0,5	1	1
Економічний вимір	0,5	1	0,5
Екологічний вимір	1	1	1

Соціальний вимір	0,5	1	0
Безпека життя	1	1	1
Злочинність	1	1	1
Зниження тривалості життя	0,5	1	0,5
Корупція	0	1	0
Соціальна нерівність	0	1	0
Рівень безробіття	1	1	1
Зношеність технологічної інфраструктури	0,5	1	1
Техногенна небезпека	0,5	1	0
Зниження добробуту населення	1	1	1
Забруднення довкілля	1	1	0,5
Дефіцит водних ресурсів	1	1	1
Смертність населення	1	1	1
Вплив на змінювання клімату	1	1	0,5

Обрахування середніх значень для складових індикаторів за формулою

$$T = \frac{\sum_{n=1}^n y}{n}, \quad (4.31)$$

де $T = \{Isd, Csl, Cql\}$, y – вірогідність наявності просторового тренду даних, n – кількість складових індикаторів тренду. В результаті було отримано значення для індексу сталого розвитку, безпеки та якості життя, відображені в таблиці 3.

Таблиця 4.3.

Порівняльний аналіз вірогідності наявності просторового тренду

Назва індексу	Середній вклад тренду при картографічному аналізі	Середній вклад тренду при поліномі 2 порядку
<i>Isd</i>	0,72	0,66
<i>Csl</i>	0,70	0,625
<i>Cql</i>	0,66	0,50

З таблиці 4.3 можна побачити, що індекс сталого розвитку має найбільші значення присутності тренду, а значення для індексу безпеки життя перевищують відповідні показники для індексу якості життя за обома видами аналізу, що вказує на переважання просторових тенденцій у даних індексу безпеки життя. В структурі індексу безпеки життя найбільший внесок до складових тренду роблять такі соціальні загрози як злочинність, рівень безробіття, зниження добробуту населення та смертність населення; економічний фактор зношеності технологічної інфраструктури та екологічні фактори дефіциту водних ресурсів, забруднення довкілля та впливу на зміну клімату. Наявність геопросторових трендів в цих даних пов'язана зі зміною природних умов по території України та диспропорціями регіонального розвитку в економічному та соціальному аспектах [137].

Використання точних методів інтерполяції дозволяє перейти від моделювання трендових поверхонь, які вказують на тенденції в зміні складових сталого розвитку, до моделювання квазістатистичних поверхонь розподілу складових сталого розвитку. Подібні поверхні дозволяють проводити виміри за показниками складових сталого розвитку на локальному рівні регіонального розвитку та розраховувати моделі на рівні адміністративних одиниць нижчого рівня ніж обласний. Точність таких розрахунків обмежується невеликою кількістю доступних точок вимірів, яка у поєднанні з багатовекторним простором утворює складну систему управління процесом сталого розвитку.

Порівняльний аналіз застосування детермінованих та стохастичних методів інтерполяції показників якості та безпеки життя для побудови квазістатистичних поверхонь показав переважання методів стохастичного моделювання. Для

порівняльного моделювання було використано метод ординарного крігінгу та обернено-зваженої відстані (IDW). Результати моделювання відображено на рис. В результаті моделювання з використанням методу ординарного крігінгу було отримано модель плавним розподілом значень, яка враховує вплив декількох значень в кожній точці. Підбір варіограмної моделі було оптимізовано з використанням сферичної моделі варіограми, яка дозволила мінімізувати дисперсію значень моделі.

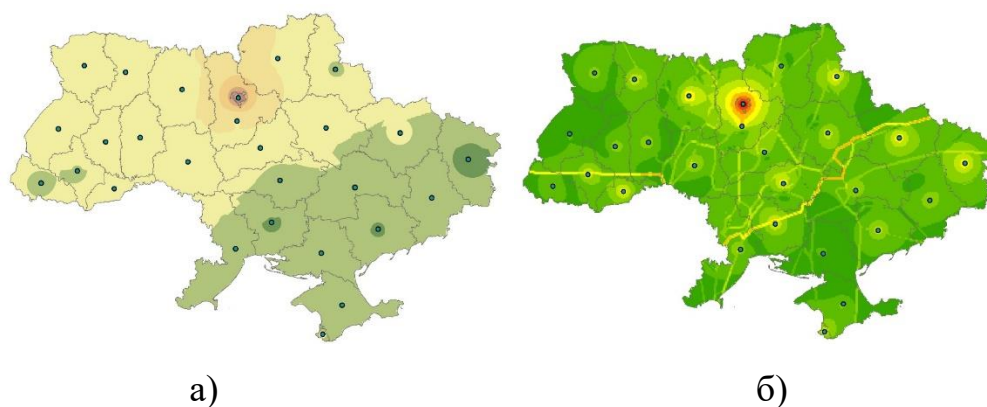


Рис. 4.15 Оцінка інтерполяційного розподілу виміру якості життя населення методами а – ординарного крігінгу; б – обернено зваженої відстані.

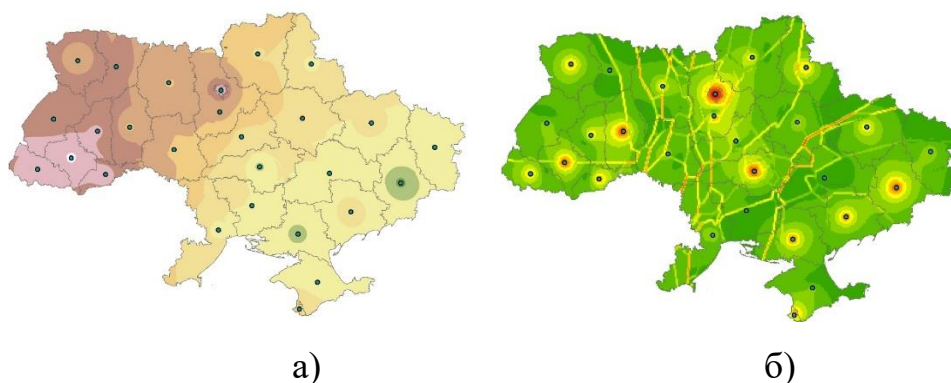


Рис. 4.16. Оцінка інтерполяційного розподілу виміру безпеки життя населення методами а – ординарного крігінгу; б – обернено зваженої відстані.

Результати отримані за допомогою методу обернено-зваженої відстані відповідають кусково-лінійним розривним функціям розподілу, які гірше описують точність методу, оскільки не враховують частину сусідніх близьких значень.

4.6. Аналіз якості методів інтерполяції даних з використанням мови R

Для визначення керуючих параметрів функції та оптимізації варіограмної моделі було проведено порівняльний аналіз детермінованих та стохастичних методів інтерполяції на основі моделювання даних геологічного знімання рівня вмісту миш'яку у підземних водах на території України. Якість підземних вод є важливою складовою управління територіальним розвитком на засадах сталого розвитку в умовах кліматичних змін та зменшення водного балансу на значній частині території України. Дослідження ґрунтується на пробах ґрунтових вод (6040 проб), в яких було виміряно вміст небезпечного для людини хімічного елементу миш'яку.

Для проведення геостатистичного аналізу, було використано мову програмування R. Це мова програмування і програмне середовище для статистичних обчислень, аналізу та представлення даних в графічному вигляді. R розповсюджується безкоштовно за ліцензією GNU General Public License у вигляді вільно доступного вихідного коду або відкомпільованих бінарних версій більшості операційних систем: Linux, FreeBSD, Microsoft Windows, Mac OS X, Solaris. R використовує текстовий інтерфейс, однак існують різні графічні інтерфейси користувача [199].

R має значні можливості для здійснення статистичних аналізів, включаючи лінійну і нелінійну регресію, класичні статистичні тести, аналіз часових рядів (серій), кластерний аналіз і багато іншого. R легко розширюється завдяки використанню додаткових функцій і пакетів доступних на сайті Comprehensive R Archive Network (CRAN). Більша частина стандартних функцій R, написана мовою R, однак існує можливість підключати код написаний C, C++, або Фортраном. Також за допомогою програмного коду на C або Java можна безпосередньо маніпулювати R об'єктами [199].

Для встановлення, необхідно скачати інсталятор за посиланням: <https://cloud.r-project.org/>. Після скачування, встановлюємо згідно вказівок інсталятора.

Далі необхідно встановити середовище розробки для мови програмування R. Для цього було обрано RStudio. Це середовище написано мовою програмування C++ і використовує фреймворк Qt для графічного інтерфейсу користувача. Для завантаження потрібно перейти за посиланням <https://www.rstudio.com/>. Далі потрібно діяти згідно вказівкам інсталятора.

Встановими необхідних бібліотек. Спершу встановимо основну бібліотеку, яка містить основні класи та методи для роботи з просторовими даними. Це робиться наступним чином:

```
install.packages("sp")
```

Встановимо бібліотеку "foreign", яка знадобиться для зчитування даних у форматі .dbf.

```
install.packages("foreign")
```

Встановимо бібліотеку "maps", яка дозволяє створювати карти, в даному випадку нам буде потрібна карта України.

```
install.packages("maps")
```

Встановимо бібліотеку "mapdata", яка містить деякі додаткові дані з карт.

```
install.packages("mapdata")
```

Встановимо бібліотеку "gstat", яка призначена для просторового геостатистичного моделювання, прогнозування та симуляції.

```
install.packages("gstat")
```

Встановимо бібліотеку "ggplot2" для візуалізації.

```
install.packages("ggplot2")
```

Після встановлення всіх цих бібліотек, приєднаємо їх до середовища.

```
library("foreign")
```

```
library("maps")
```

```
## Warning: package 'maps' was built under R version 3.3.1
```

```
##
```

```
## # maps v3.1: updated 'world': all lakes moved to separate new
```

```
#
```

```
## # 'lakes' database. Type '?world' or 'news(package="maps")'.
```

```
#
```

```
library("mapdata")
```

```
## Warning: package 'mapdata' was built under R version 3.3.1
library(ggplot2)
library(sp)
## Warning: package 'sp' was built under R version 3.3.1
library(gstat)
## Warning: package 'gstat' was built under R version 3.3.1
На цьому встановлення необхідних програм і пакунків завершено.
```

Підготовка та огляд даних

Дані було надано Світовим центром даних з геоінформатики та сталого розвитку. Формат даних - dbf. Зчитаємо дані з файлу "Point_2.dbf" у змінну x.

```
x <- read.dbf("Point_2.dbf")
```

Дані зчитали у форматі data.frame. Тепер дослідимо які маємо стовпці, їх назви, та типи значень. Зробимо це за допомогою команди str, завдяки якій можна відобразити структуру об'єкта в середовищі R.

```
str(x)
## 'data.frame':    6553 obs. of  42 variables:
## $ GRE      : int  52 52 52 52 52 52 52 52 52 52 ...
## $ GODO     : int  82 82 82 82 82 83 83 83 83 82 ...
## $ TYPOBJ   : int   8 8 8 8 8 8 8 8 8 8 ...
## $ NOBJ     : Factor w/ 0 levels: NA NA NA NA NA NA NA ...
## $ SUBSTAN  : int  20 20 20 20 20 20 20 20 20 20 ...
## $ NPRB     : Factor w/ 6553 levels "10001","10002",...: 5352 5374 5469 5491 6505 282 391 455 1097 5144 ...
## $ NP       : int  639 641 650 652 755 1110 1126 1193 2201 620 ...
## $ LABOR    : int   1 1 1 1 1 1 1 1 1 1 ...
## $ LABNO    : int  622 622 622 622 622 433 433 498 1001 622 ...
## $ DBXN     : Date, format: "1982-06-30" "1982-06-30" ...
## $ U1       : num  1e-05 1e-05 1e-05 1e-05 1e-05 1e-05 1e-05 1e-05 1e-05 1e-05 ...
## $ BK       : num  0.000125 0.000125 0.000125 0.000125 0.000125 0.000125 0.000125 0.000125 0.000125 0.000125 ...
## $ AS1      : num  0.004 0 0 0 0 0 0 0.002 0 0 ...
## $ F1       : num  0 0 0 0 0.15 0.13 0.11 0.19 0 0 ...
## $ HCO3     : num  -1 -1 -1 -1 -1 -1 -1 85.4 30.5 -1 ...
## $ UOBT     : num  0 0 0 0 0 0 0 0 0 0 ...
## $ REGION   : Factor w/ 19 levels "49-34","KAP",...: 17 3 4 4 17 6 6 17 17 17 ...
## $ GRE_2    : int  52 52 52 52 52 52 52 52 52 52 ...
## $ TYPOBJ_2 : int   8 8 8 8 8 8 8 8 8 8 ...
## $ SUBSTAN_2: int  20 20 20 20 20 20 20 20 20 20 ...
## $ NPRB_2   : Factor w/ 6553 levels "10001","10002",...: 5352 5374 5469 5491 6505 282 391 455 1097 5144 ...
## $ NPRB1    : int  639 641 650 652 755 1110 1126 1193 2201 620 ...
## $ DO       : int   9 10 11 11 22 25 27 20 22 8 ...
## $ MO       : int   6 6 6 6 6 4 4 5 9 6 ...
## $ GODO_2   : int  82 82 82 82 82 83 83 83 83 82 ...
## $ GLUB     : num  0 0 0 0 0 0 0 0 0 0 ...
## $ OBEM     : num  0 0 0 0 0 0 0 0 0 0 ...
## $ SG       : int  51 51 51 51 51 50 51 50 50 51 ...
## $ SM       : int  31 37 34 39 8 55 14 42 42 2 ...
## $ SS       : num  50 44 40 20 31 0 50 31 20 32 ...
## $ DG       : int  27 27 27 27 28 30 30 28 27 27 ...
## $ DM       : int  25 29 7 5 51 26 27 36 35 41 ...
## $ DS       : num  20 17 40 11 36 0 52 32 33 33 ...
## $ X        : num  5711332 5722301 5716510 5725159 5669657 ...
## $ Y        : num  5529300 5533796 5508858 5505979 5630166 ...
## $ Z        : num  0 0 0 0 0 0 0 0 0 0 ...
## $ NL       : Factor w/ 1716 levels "L-34-012-B","L-35-001-A",...: 554 554 553 553 607 1100 1063 648 641 598 ...
## $ LANSH    : int   0 0 0 0 0 0 0 0 0 0 ...
## $ REGI     : int   0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ BLK      : int  0 0 0 0 0 0 0 0 0 ...
## $ SH       : num  51.5 51.6 51.6 51.6 51.1 ...
## $ DH       : num  27.4 27.5 27.1 27.1 28.9 ...
## - attr(*, "data_types")= chr  "N" "N" "N" "C" ...
```

Тепер поглянемо більш детально статистичні розподіл кожного показника. За допомогою команди `summary` можна побачити інтервал значень кожного стовпця та квартилі.

summary(x)

```
##      GRE      GODO      TYPOBJ      NOBJ      SUBSTAN
## Min.   : 0.00   Min.   : 5.00   Min.   : 8   NA's:6553   Min.   :20
## 1st Qu.:52.00   1st Qu.:84.00   1st Qu.: 8   1st Qu.:20
## Median :52.00   Median :85.00   Median : 8   Median :20
## Mean   :51.79   Mean   :85.54   Mean   : 8   Mean   :20
## 3rd Qu.:52.00   3rd Qu.:87.00   3rd Qu.: 8   3rd Qu.:20
## Max.   :52.00   Max.   :91.00   Max.   : 8   Max.   :20
##
##      NPRB      NP      LABOR      LABNO
## 10001 : 1   Min.   : 56   Min.   :1   Min.   : 381
## 10002 : 1   1st Qu.: 2545   1st Qu.:1   1st Qu.: 549
## 10003 : 1   Median : 4269   Median :1   Median : 655
## 10004 : 1   Mean   : 4339   Mean   :1   Mean   : 669
## 10005 : 1   3rd Qu.: 6048   3rd Qu.:1   3rd Qu.: 785
## 10006 : 1   Max.   :11201   Max.   :1   Max.   :1164
## (Other):6547
##      DBXN      U1      BK
## Min.   :1981-08-31   Min.   :0.000010   Min.   :0.000125
## 1st Qu.:1984-07-31   1st Qu.:0.000650   1st Qu.:0.008125
## Median :1985-09-09   Median :0.001600   Median :0.020000
## Mean   :1986-02-16   Mean   :0.003868   Mean   :0.048349
## 3rd Qu.:1987-08-24   3rd Qu.:0.003790   3rd Qu.:0.047375
## Max.   :1991-09-24   Max.   :0.145000   Max.   :1.812500
##
##      AS1      F1      HCO3      UOBT
## Min.   :-9.990000   Min.   :-9.9000   Min.   : -1.0   Min.   :0
## 1st Qu.: 0.000000   1st Qu.: 0.1500   1st Qu.: 207.5   1st Qu.:0
## Median : 0.002000   Median : 0.3400   Median : 305.1   Median :0
## Mean   :-0.001088   Mean   :-0.0318   Mean   : 306.6   Mean   :0
## 3rd Qu.: 0.006000   3rd Qu.: 0.5500   3rd Qu.: 390.5   3rd Qu.:0
## Max.   : 0.040000   Max.   : 9.9990   Max.   : 9999.9   Max.   :0
##
##      REGION      GRE_2      TYPOBJ_2      SUBSTAN_2      NPRB_2
## УЩ-ЦЕНТР :1031   Min.   :47.0   Min.   : 8   Min.   :20   10001 : 1
## ВП       : 890   1st Qu.:52.0   1st Qu.: 8   1st Qu.:20   10002 : 1
## ДДВ-КИЕВ : 717   Median :52.0   Median : 8   Median :20   10003 : 1
## ДДВ-ЦЕНТР: 592   Mean   :51.8   Mean   : 8   Mean   :20   10004 : 1
## ПЧ-ВОСТОК: 501   3rd Qu.:52.0   3rd Qu.: 8   3rd Qu.:20   10005 : 1
## (Other) :2722   Max.   :52.0   Max.   : 8   Max.   :20   10006 : 1
## NA's     :100   (Other):6547
##      NPRB1      DO      MO      GODO_2
## Min.   : 56   Min.   : 0.00   Min.   : 0.00   Min.   : 5.00
## 1st Qu.: 2545   1st Qu.: 7.00   1st Qu.: 6.00   1st Qu.:84.00
## Median : 4269   Median :14.00   Median : 7.00   Median :85.00
## Mean   : 4339   Mean   :13.91   Mean   : 6.41   Mean   :85.54
## 3rd Qu.: 6048   3rd Qu.:20.00   3rd Qu.: 8.00   3rd Qu.:87.00
## Max.   :11201   Max.   :31.00   Max.   :10.00   Max.   :91.00
##
##      GLUB      OBEM      SG      SM      SS
## Min.   :0   Min.   :0   Min.   :44.00   Min.   : 0.00   Min.   : 0.00
## 1st Qu.:0   1st Qu.:0   1st Qu.:48.00   1st Qu.:14.00   1st Qu.:12.00
## Median :0   Median :0   Median :49.00   Median :28.00   Median :28.00
## Mean   :0   Mean   :0   Mean   :48.59   Mean   :28.72   Mean   :27.73
## 3rd Qu.:0   3rd Qu.:0   3rd Qu.:50.00   3rd Qu.:43.00   3rd Qu.:43.71
## Max.   :0   Max.   :0   Max.   :52.00   Max.   :59.00   Max.   :73.00
##
##      DG      DM      DS      X
## Min.   :22.00   Min.   : 0.0   Min.   : 0.00   Min.   :4925879
## 1st Qu.:28.00   1st Qu.:14.0   1st Qu.:11.00   1st Qu.:5347245
## Median :31.00   Median :29.0   Median :27.00   Median :5446196
## Mean   :30.75   Mean   :29.1   Mean   :26.66   Mean   :5439993
## 3rd Qu.:34.00   3rd Qu.:44.0   3rd Qu.:42.00   3rd Qu.:5563060
## Max.   :40.00   Max.   :77.0   Max.   :59.52   Max.   :5800987
```

```
##
##      Y              Z              NL              LANSH
## Min.   :4588050   Min.   : -1.00000   M-35- 57-B: 42   Min.   :0
## 1st Qu.:5574900   1st Qu.: 0.00000   M-35- 69-A: 41   1st Qu.:0
## Median :6376402   Median : 0.00000   M-36-099-B: 25   Median :0
## Mean   :6212549   Mean   : -0.01526   M-36-087-Г: 24   Mean   :0
## 3rd Qu.:6605234   3rd Qu.: 0.00000   M-35-120-Г: 23   3rd Qu.:0
## Max.   :7582968   Max.   : 0.00000   M-35- 56-B: 22   Max.   :0
##                                     (Other) :6376
##      REGI          BLK          SH          DH
## Min.   :0   Min.   :0   Min.   :44.45   Min.   :22.18
## 1st Qu.:0   1st Qu.:0   1st Qu.:48.23   1st Qu.:28.03
## Median :0   Median :0   Median :49.12   Median :31.30
## Mean   :0   Mean   :0   Mean   :49.07   Mean   :31.23
## 3rd Qu.:0   3rd Qu.:0   3rd Qu.:50.18   3rd Qu.:34.40
## Max.   :0   Max.   :0   Max.   :52.33   Max.   :40.13
##
```

Нас цікавить стовпець AS1(арсен), що характеризує кількісний вміст миш'яку. Як бачимо, мінімальне значення становить -9.9, що свідчить про те, що в базі наявні викиди. Не акцентуватимемо увагу на кількісних значеннях інших ресурсів, залишимо в нашому датафреймі лише ті записи, в яких кількісне значення миш'яку не менше нуля.

```
x<-x[x$AS1>=0,]
```

Стовпці SH та DH характеризують широту та довготу в якій проводилося дане дослідження. Як бачимо за інтервалом значень, всі досліджувані точки знаходяться в межах України. Для проведення більш ефективної інтерполяції, вилучимо дослідження, що проводились в однакових точках, тобто дублікати.

```
x<-x[!(duplicated(x[c("DH", "SH")]) | duplicated(x[c("DH", "SH")],
fromLast = TRUE)), ]
```

Побудуємо гістограму для миш'яку (рис. 4.34).

```
ggplot(x, aes(x=AS1))+geom_histogram(bins=70)
```

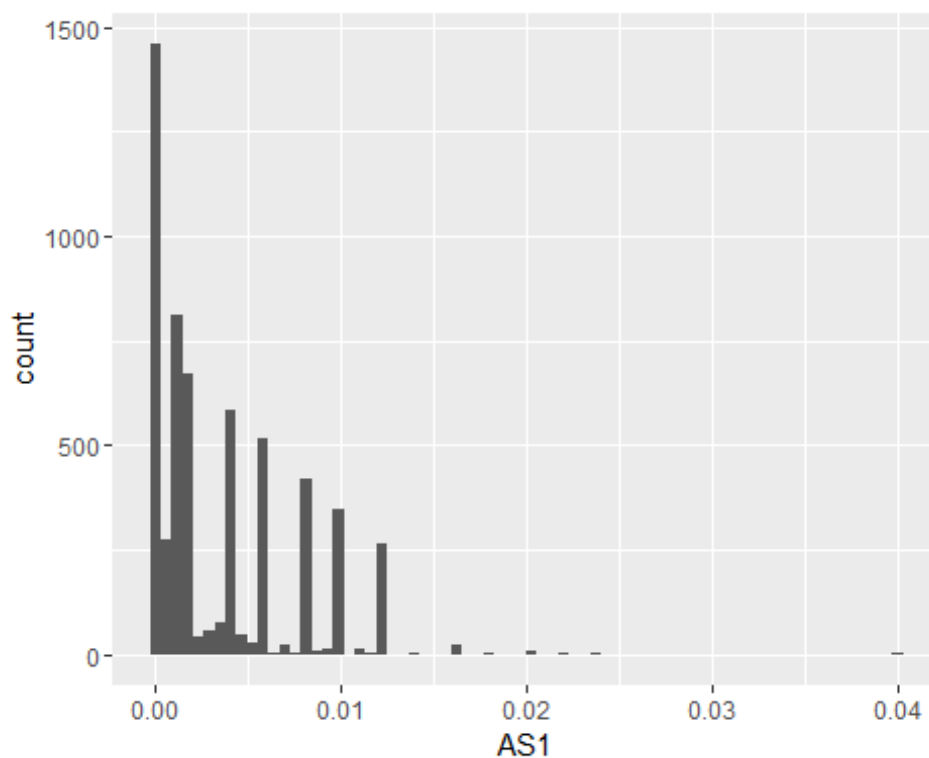


Рис. 4.17. Гістограма розподілу значень миш'яку у підземних водах України у сукупності проб (за рівнями)

Побудуємо гістограму, яка описує середні значення миш'яку по регіонах (4.18).

```
x.sum <- aggregate(x["AS1"], x["REGION"], FUN=mean)
ggplot(x.sum, aes(x=REGION, y=AS1)) +
  geom_bar(stat="identity")+theme(text = element_text(size=10),
    axis.text.x = element_text(angle=90, vjust=1))
```

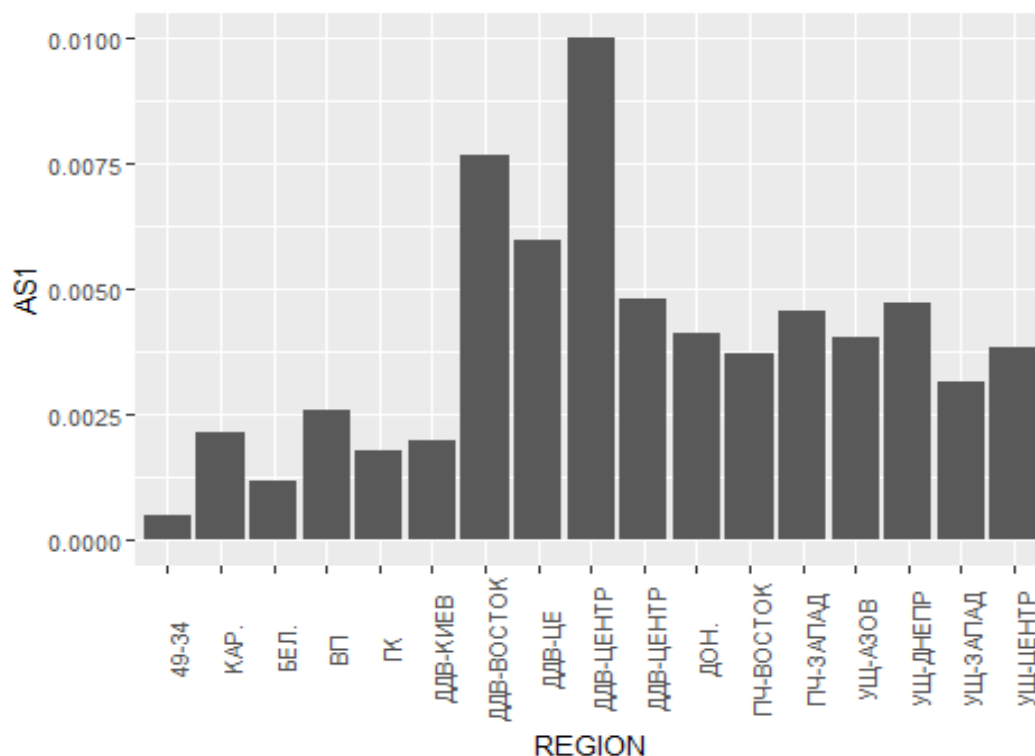


Рис. 4.18. Гістограма розподілу значень миш'яку за геологічними провінціями у підземних водах

Тепер запишемо карту України з просторових даних в змінну `Ukraine`.

```
Ukraine<-map(regions='Ukraine',
             xlim=c(22,41), ylim=c(44, 53),
             fill=TRUE,plot=FALSE)
```

Визначення приблизних меж, в яких знаходиться Україна, дає факт, що поточна система координат є у вигляді широта та довгота. Візуалізуємо на карті всі точки, в яких проводилась геологічна зйомка (рис. 4.19).

```
ggplot() +
  geom_path(data = Ukraine, aes(x = long, y = lat, group =
group)) +
  geom_point(data = x, aes(x = DH, y = SH), color = "red")
```

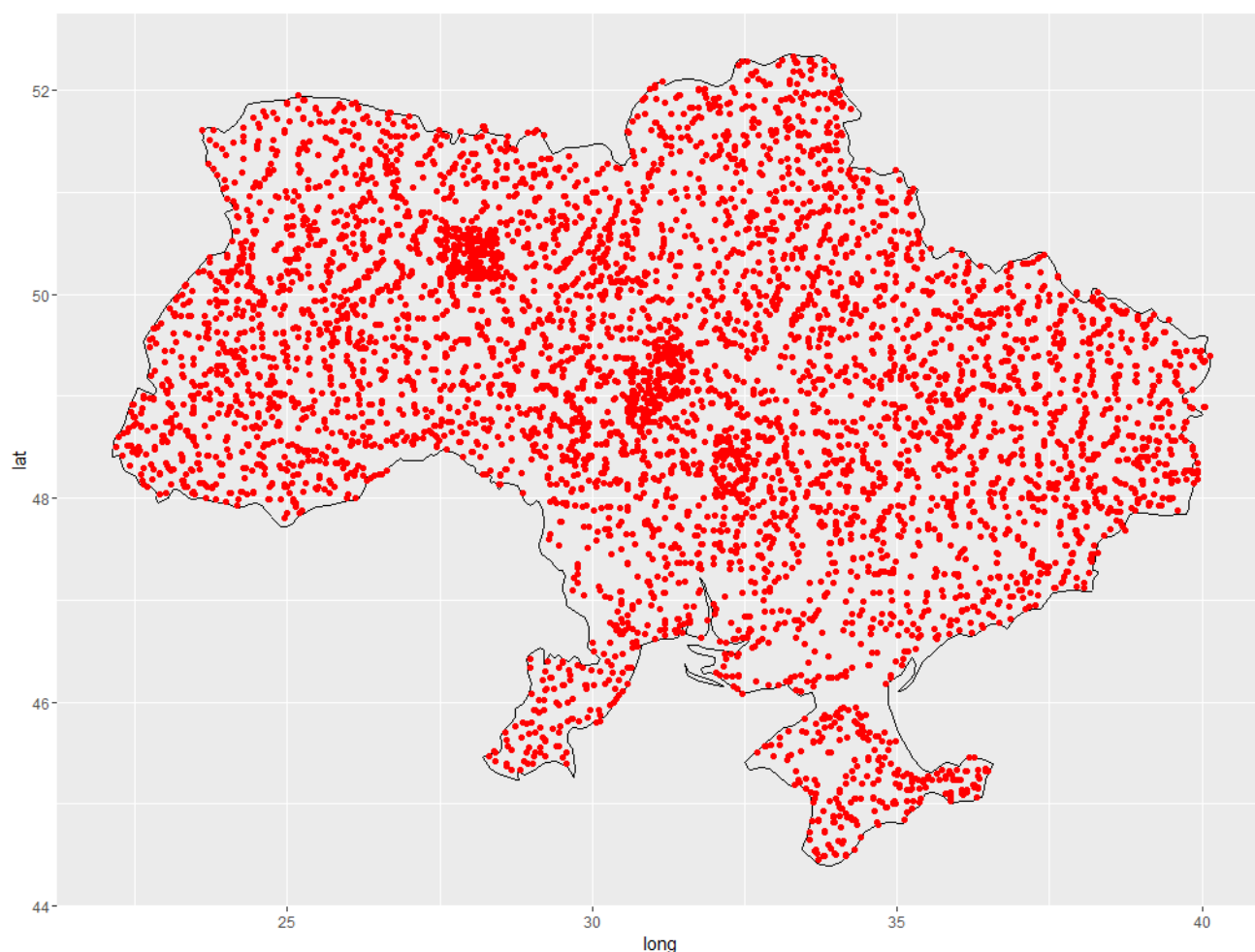


Рис. 4.19. Точки відбору проб підземних вод

Тепер зобразимо у "бульбашкового" графіку, який характеризує кількісно вміст миш'яку в кожній досліджуваній точці. При побудові приберемо всі нульові значення, щоб вони не заважали (4.20).

```
ggplot() +  
  geom_path(data = Ukraine, aes(x = long, y = lat, group =  
group)) +  
  geom_point(data = x[x$AS1>0,], aes(x = DH, y = SH, size =  
AS1), color = "red")
```

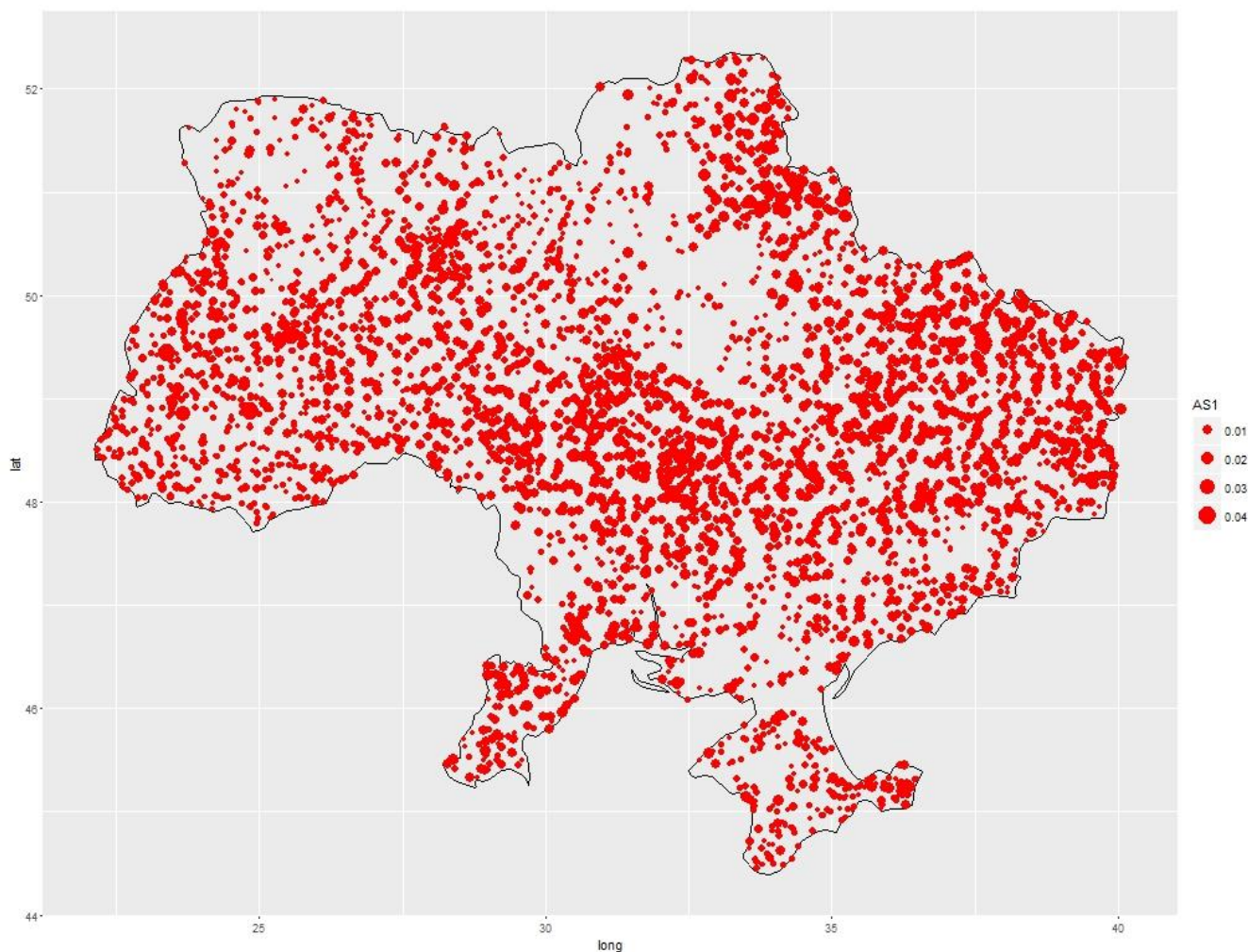



Рис. 4.20. Розподіл миш'яку у пробах води за географічним положенням

Для подальшого аналізу, перетворимо наш датасет в тип `SpatialPointsDataFrame`, виділивши туди як дані стовпчику арсену, а в якості координат довготу і широту.

```
spdf<-SpatialPointsDataFrame(cbind(x$DH, x$SH),
as.data.frame(AS1),proj4string=CRS("+proj=longlat
+datum=WGS84"))
```

Створимо прямокутник класу `SpatialPixels`, який містить в собі Україну. Для дискретизації оберемо прямокутники по 0.05 градуси довготи і широти. В подальшому на ньому проводитимемо інтерполяцію.

```
x.range <- as.integer(c(21.0,42.0))
y.range <- as.integer(c(44.0,53.0))
grd <- expand.grid(x=seq(from=x.range[1], to=x.range[2],
by=0.05), y=seq(from=y.range[1], to=y.range[2], by=0.05))
coordinates(grd) <- ~ x+y
gridded(grd) <- TRUE
proj4string(grd)<-CRS("+proj=longlat +datum=WGS84")
```

Виріжемо з цього прямокутника територію України. На ній проводитимемо інтерполяцію.

```
Ukrainena<-Ukraine
Ukrainena$x<-Ukraine$x[!is.na(Ukraine$x)]
Ukrainena$y<-Ukraine$y[!is.na(Ukraine$y)]
inner<-
point.in.polygon(grd@coords[,1],grd@coords[,2],Ukrainena$x,Ukrainena$y)
inner<-as.logical(inner)
grd<-grd[inner,]
```

На цьому первинна обробка та огляд даних завершено, переходимо до інтерполяційних методів.

Проведення інтерполяції

Метод обернено зважених відстаней

Інтерполяція за методом зважених відстаней (IDW) використовує припущення, що об'єкти розташовані ближче до другого в більшій мірі схожі, ніж віддалені одне від одного. Щоб знайти значення в якій-небудь точці, метод ОЗВ використовує опорні точки, що розташовані в околі шуканої.

Якщо вибірки з ресурсами відносно рівномірно розподілені, і характеристики поверхні особливо не змінюються в різних частинах ландшафту, можна з достатньою точністю інтерполювати значення поверхні на основі значень поруч розташованих точках. Щоб врахувати різну віддаленість точок від шуканої, значенням опорних точок, розташованих ближче до неї, присвоюється більша вага.

Це основа методу інтерполяції, відомого як метод обернено зважених відстаней – Inverse Distance Weighting (IDW). Як впливає з назви, вага значення зменшується по мірі збільшення відстані від шуканої точки.

Ці опорні точки будуть створювати більший вплив на інтерпольоване значення, ніж ті, які від неї віддалені на значну відстань. Таким чином, метод ОЗВ припускає, що кожна опорна точка створює локальний вплив, який зменшується з відстанню [199].

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

Де $\hat{Z}(s_0)$ – шукане значення для точки s_0

Де λ_i – ваги, присвоєні кожній опорній точці, з числа тих, що використовуватимуться в обчисленнях. Ці ваги зменшуються з відстанню.

Де $Z(s_i)$ – виміряне значення в точці s_i .

N – число опорних точок, що знаходяться в околі шуканої точки і тих, що використовуються для обчислень.

Ваги визначаються за наступною формулою:

$$\lambda_i = d_{i0}^{-p} / \sum_{i=1}^N d_{i0}^{-p}$$

Де d_{i0} - це відстань між шуканою точкою s_0 і i -тою опорною точкою s_i .

Зі збільшенням відстані, вага зменшується за рахунок степені p .

Параметр степені p впливає на присвоєння ваги опорним точкам; це означає, що в міру того як збільшується відстань між опорними точками і шуканою, вплив (або вага), яка опорна точка буде створювати на шукану, зменшується експоненційно.

Сума ваг опорних точок, які будуть використовуватися при інтерполяції, повинна бути рівна 1.

$$\sum_{i=1}^N \lambda_i = 1$$

Подивимось як в залежності від степені p буде змінюватися отримана поверхня. Спочатку за допомогою команди `idw` проведемо інтерполяцію методом обернено зважених відстаней (рис. 4.21).

```
idw.out<-idw(AS1 ~ 1, locations=spdf, newdata=grd, idp=1.0)
## [inverse distance weighted interpolation]
```

Тепер переведемо отриманий результат з класу `SpatialPixelsDataFrame` до звичайного `data.frame` і візуалізуємо результат за допомогою `ggplot`.

```
idw.output <- as.data.frame(idw.out)
names(idw.output)[1:3] <- c("long", "lat", "var1.pred")
ggplot() + geom_tile(data = idw.output, aes(x = long, y = lat,
```

```
fill = var1.pred)) + scale_fill_gradient2(midpoint =
median(idw.out@data$var1.pred, na.rm = T)) +
  geom_path(data = Ukraine, aes(long, lat, group = group), colour =
"grey40", size=1) + labs(fill = "Миш'як", title = "Розподіл миш'яку по Україні")
```

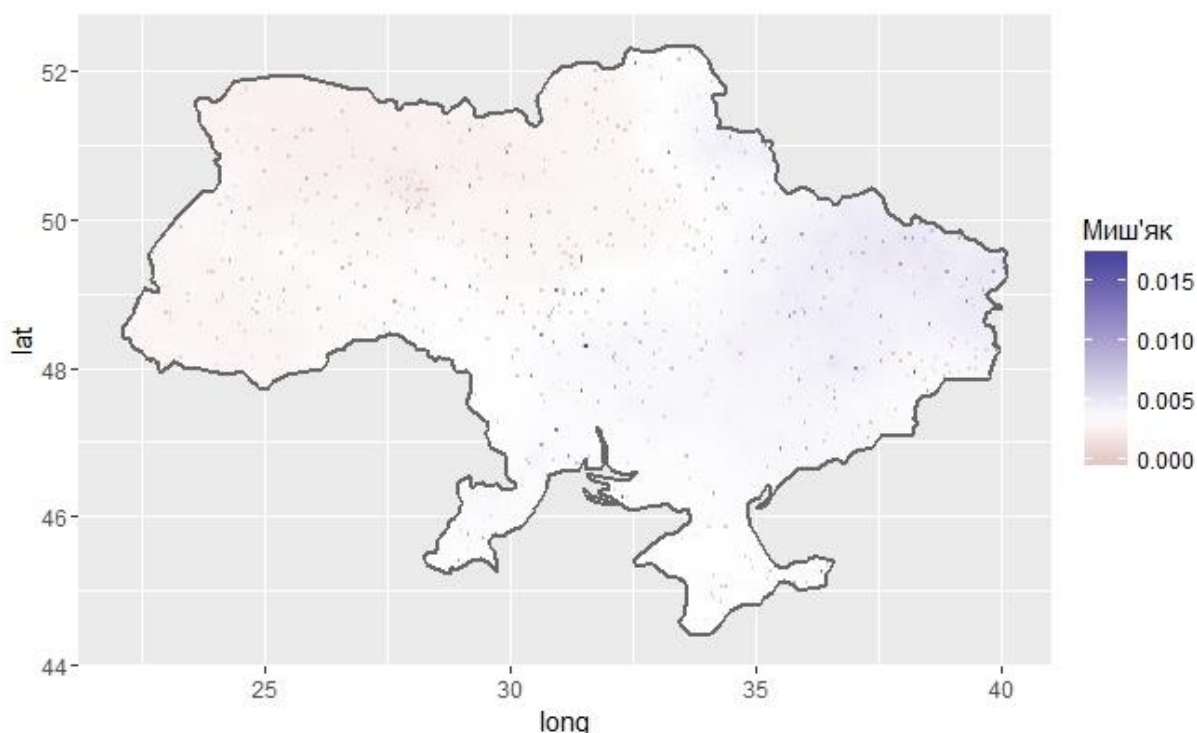


Рис. 4.21. Інтерполяція значень методом обернено зваженої відстані (ступінь ваги дорівнює 1)

Тепер аналогічно поглянемо, який буде результат при великих ступенях.

При ступені 2 (рис. 4.22):

```
idw.out<-idw(AS1 ~ 1, locations=spdf, newdata=grd, idp=2.0)
idw.output=as.data.frame(idw.out)
names(idw.output)[1:3]<-c("long","lat","var1.pred")
ggplot() + geom_tile(data = idw.output, aes(x = long, y = lat,
fill = var1.pred)) + scale_fill_gradient2(midpoint =
median(idw.out@data$var1.pred, na.rm = T)) +
  geom_path(data = Ukraine, aes(long, lat, group = group), colour =
"grey40", size=1) + labs(fill = "Миш'як", title = "Розподіл миш'яку по Україні")
```

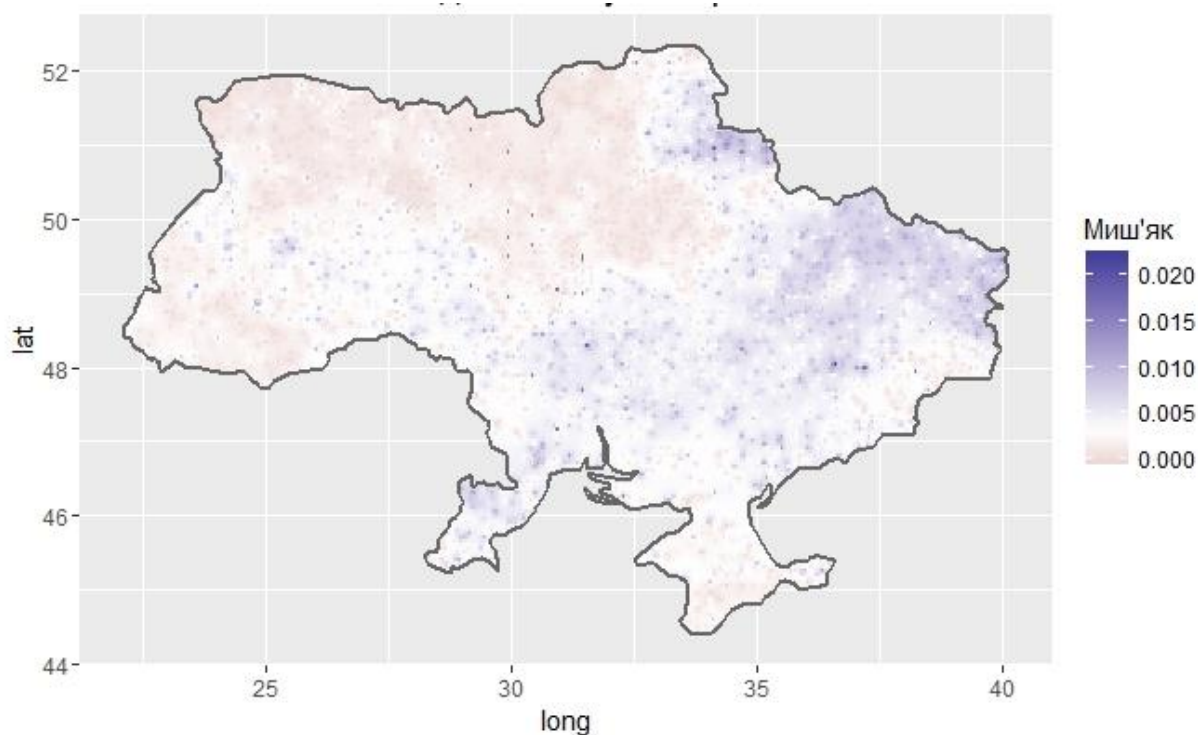


Рис. 4.22. Інтерполяція значень методом обернено зваженої відстані (ступінь ваги дорівнює 2)

При ступені 3 (рис. 4.23):

```
idw.out<-idw(AS1 ~ 1, locations=spdf, newdata=grd, idp=3.0)
idw.output=as.data.frame(idw.out)
names(idw.output)[1:3]<-c("long","lat","var1.pred")
ggplot() + geom_tile(data = idw.output, aes(x = long, y = lat,
fill = var1.pred)) + scale_fill_gradient2(midpoint =
median(idw.out@data$var1.pred,na.rm = T)) +
  geom_path(data = Ukraine, aes(long, lat, group = group), colour =
"grey40",size=1) + labs(fill = "Миш'як", title = "Розподіл миш'яку по Україні")
```

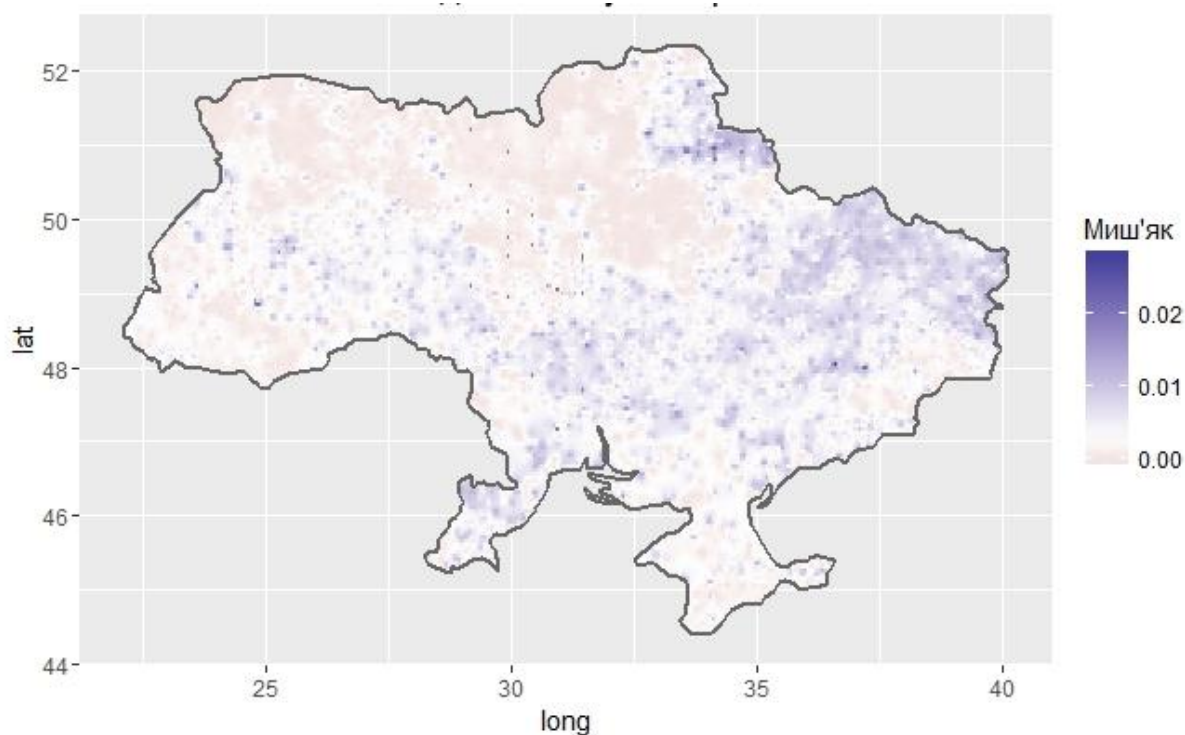


Рис. 4.23. Інтерполяція значень методом обернено зваженої відстані (ступінь ваги дорівнює 3)

Метод крігінга

Інтерполяція методом ОЗВ відноситься до детермінованих методів інтерполяції, оскільки він напряму базується на вимірних значеннях, що потрапляють в окіл інтерпольованої точки, і на заданих математичних формулах, які визначають гладкість поверхні. Друга група методів інтерполяції складається з геостатистичних методів, таких як крігінг, які базуються на статистичних моделях, що включають аналіз автокореляції (статистичних відношень між вимірними очками). В результаті цього геостатистичні методи не тільки мають можливість створювати поверхні прогнозованих значень, а також надають деякі вимірювання достовірності чи точності прогнозованих значень [198].

При крігінгу припускається, що відстань чи напрям між опорними точками відображає поверхневу кореляцію, яка може використовуватися для пояснення змін на поверхні. Крігінг використовує математичні функції для певної кількості точок або всіх точок в межах заданого радіусу, щоб визначити вихідне значення для всіх напрямів. Крігінг – процес покроковий; він включає пошуковий

статистичний аналіз даних, моделювання варіограми, створення поверхні і (додатково) вивчення поверхні дисперсії [297].

Крігінг аналогічний ОЗВ в тому, що він зважує оточуючі вимірювані значення, щоб отримати передбачення для невимірюваного місця. Основна формула для цих двох інструментів інтерполяцій формується як зважена сума даних:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

Де $\hat{Z}(s_0)$ – шукане значення для точки s_0

Де λ_i – ваги, присвоєні кожній опорній точці, з числа тих, що використовуватимуться в обчисленнях. Ці ваги зменшуються з відстанню.

Де $Z(s_i)$ – виміряне значення в точці s_i .

N – число опорних точок, що знаходяться в околі шуканої точки і тих, що використовуються для обчислень.

В ОЗВ вага залежить лише від відстані до місцезнаходження прогнозу. Однак при використанні методу крігінгу, ваги засновані не тільки на відстані між вимірюваними точками і місцезнаходженням прогнозів, але також на загальному просторовому розташуванні вимірюваних точок. Щоб використати просторове розташування в вагах, необхідно визначити кількість просторової автокореляції.

Щоб зробити прогноз за допомогою крігінга, необхідно дві задачі:

- Розкрити залежність правил.
- Зробити прогнози.

Щоб реалізувати ці задачі, крігінг проходить через 2-кроковий процес:

1. Він створює варіограми і функції коваріації для оцінки значень статистичної залежності, які залежать від моделі автокореляції.
2. Він прогнозує невідомі значення (робить прогноз).

Установка моделі чи просторове моделювання, також відоме як структурний аналіз чи варіографія. В просторовому моделюванні структури вимірюваних точок процес починається з діаграми емпіричної варіограми, обчисленої за допомогою наступного рівняння для всіх пар місцезнаходжень, розділених відстанню h :

$$\text{Semivariogram}(\text{distance}_h) = 0.5 * \text{average}((\text{value}_i - \text{value}_j)^2)$$

Формула включає обчислення різниці між квадратом значень парних місцезнаходжень. Цей процес продовжується для кожної вимірюваної точки.

Часто кожна пара місцезнаходжень має унікальну відстань і часто існує багато пар точок. Розташування всіх пар швидко стає некерованим. Замість розташування кожної пари, вони формуються в lag bins. Наприклад, обчислюється середня варіограма для всіх пар точок, розташованих одне від одного далі ніж 40 метрів, але менш ніж 50. Емпірична варіограма – це діаграма середніх значень варіограми на осі y і відстань (або лаг) на осі x.

Просторова автокореляція дозволяє кількісно оцінити основний принцип географії: ближчі об'єкти мають більшу степінь схожості, ніж віддалені. Таким чином, пари близько розташованих місцезнаходжень (крайні зліва на осі y варіограми) повинні мати більш схожі значення (внизу на осі y в хмарі варіограми). По мірі віддалення одне від одного пари місцезнаходження (переміщення праворуч по осі x хмари варіограми) вони повинні стати більше несхожими і отримати більшу різницю квадратів (переміщення вгору по осі y хмари варіограми) [297].

Наступним кроком є встановлення моделі на точки, що формують емпіричну варіограму. Моделювання варіограми – це ключовий крок між просторовим описом і просторовим прогнозом. Основне застосування крігінга – прогноз атрибутивних значень в неопорних місцезнаходженнях. Емпірична варіограма надає інформацію про просторову автокореляцію наборів даних. Однак вона не дає інформацію для всіх можливих напрямів і відстаней. По цій причині і щоб переконатися, що прогнози крігінга мають додатні дисперсії крігінга, необхідно встановити модель – тобто неперервну функцію чи криву на емпіричну варіограму. Абстрактно кажучи, це аналогічно аналізу регресії, в якому неперервна лінія чи крива встановлюється на точкові дані. Існує багато функцій для моделювання емпіричних варіограм, наприклад: експоненційна, лінійна, кругова, сферична, гаусівська тощо.

Як відзначалось раніше, варіограма відображає просторову автокореляцію вимірюваних опорних точок. Внаслідок основного принципу географії (ближчі об'єкти схожі), вимірювані точки, що знаходяться ближче одне до одного, як

правило мають меншу різницю квадратів, ніж точки, розташовані далі. Коли кожна пара розташована після binned, через них встановлюється модель. Для опису цих моделей зазвичай використовують діапазон, поріг і самородок.

Якщо дивитись на модель варіограми, можна побачити, що на певній відстані модель вирівнюється. Відстань, де модель спочатку вирівнюється, називається діапазоном. Опорні місцезнаходження, розділені відстанями, не перевищуючими діапазон, просторово автокорельовані, в той же час місцезнаходження, що перевищують діапазон – ні (рис. 4.24).

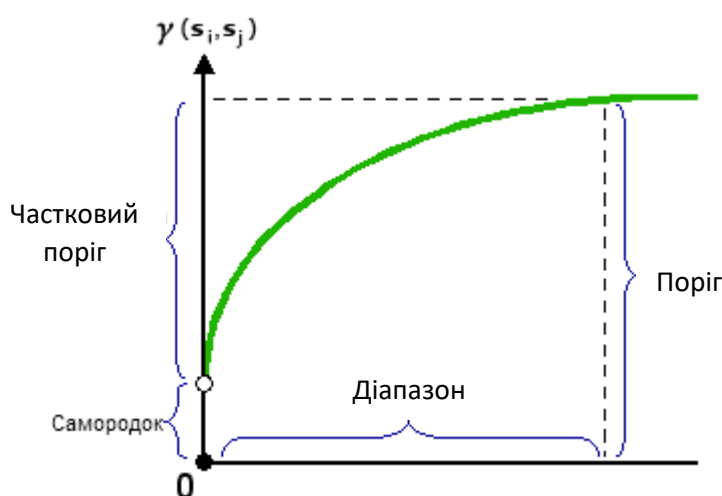


Рис. 4.24. Модель варіограми у моделі крігінгу

Значення, в якому модель варіограми досягає діапазону, називається порогом. Частковий поріг – це поріг мінус самородок.

Після того, як була знайдена залежність чи автокореляція в даних, можна зробити прогноз використовуючи встановлену модель. Таким чином, емпірична варіограма призупиняється. Тепер можна використовуючи дані, зробити прогноз. Крігінг полягає в мінімізації дисперсії похибки вимірювання, яка є функцією від вимірюваних ваг. Мінімізація цієї дисперсії зменшує середню квадратичну похибку відхилення оціненого значення від можливого. Досягається це шляхом прирівнювання до нуля першої похідною похибки щодо кожної невідомої ваги. В результаті виводиться система рівнянь, розв'язком якої є вектор ваг [297].

Доступні два методи крігінга: ординарний і універсальний.

Ординарний крігінг — найбільш загальний і широко використовуваний методі крігінга, він використовується за замовчуванням. Передбачається, що середнє значення константи невідомо. Це припущення має сенс, поки немає наукового підґрунтя відхилити його.

При універсальному крігінгу передбачається, що є домінуючий тренд в даних — наприклад, переважний вітер — і його можна моделювати детермінованою функцією, поліномом. Цей поліном отримується з вихідних вимірених точок, і автокореляція моделюється з довільних похибок. Після установки моделі на довільні похибки і до прогнозування, поліном додається назад до прогнозів, щоб дати значимі результати. Універсальний крігінг слід використовувати, тільки якщо точно відомо, що в даних є тренд, і можна дати наукове обґрунтування для його опису [298].

Побудуємо варіограму на основі наших даних (рис. 4.25):

```
semivariog<-variogram(x$AS1~1, locations=grd, data=spdf)
plot(semivariog)
```

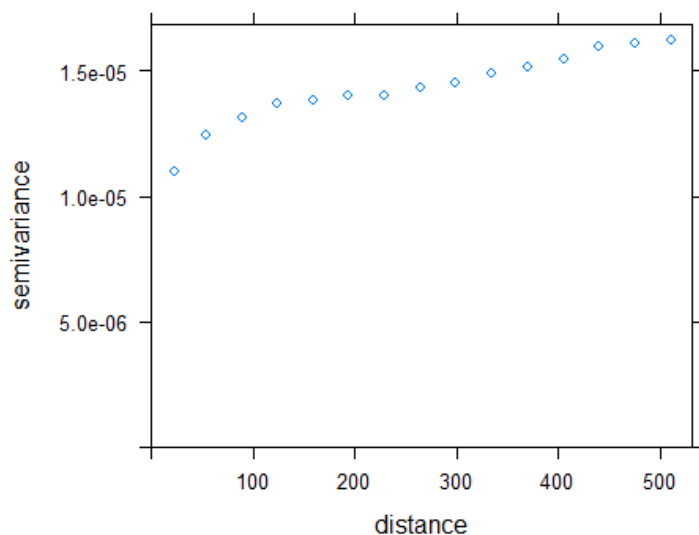


Рис. 4.25. Варіограмна модель даних про розподіл миш'яку.

Як видно з варіограми, можна встановити, що самородок дорівнює близько $1.0e-05$, поріг - $1.7e-05$, а діапазон приблизно 520. Тепер спробуємо змоделювати варіограми функціями. Подивимось на експоненційну, сферичну і гаусівську.

Експоненційна (рис. 4.42):

```
model.variog.exp<-vgm(psill=1.7e-05, model="Exp", nugget=1.0e-05, range=520)
fit.variog.exp<-fit.variogram(semivariog, model.variog.exp)
plot(semivariog, fit.variog.exp)
```

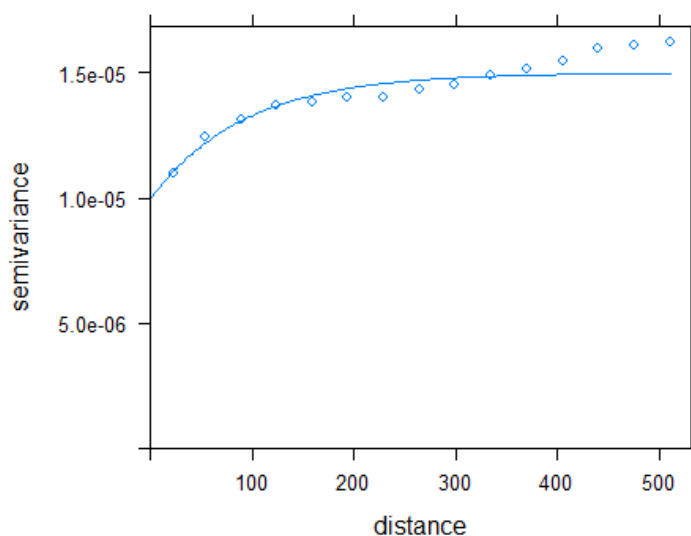


Рис. 4.42. Експоненційна варіограмна модель даних про розподіл миш'яку.

Сферична (рис. 4.26):

```
model.variog.sph<-vgm(psill=1.7e-05, model="Sph", nugget=1.0e-05, range=520)
fit.variog.sph<-fit.variogram(semivariog, model.variog.sph)
plot(semivariog, fit.variog.sph)
```

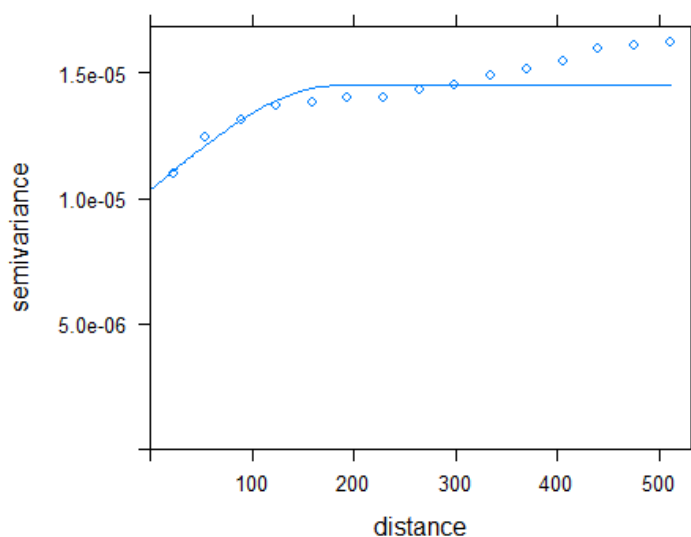


Рис. 4.26. Сферична варіограмна модель даних про розподіл миш'яку.

Гаусівська (рис. 4.27):

```
model.variog.gau<-vgm(psill=1.7e-05, model="Gau", nugget=1.0e-05, range=520)
fit.variog.gau<-fit.variogram(semivariog, model.variog.gau)
plot(semivariog, fit.variog.gau)
```

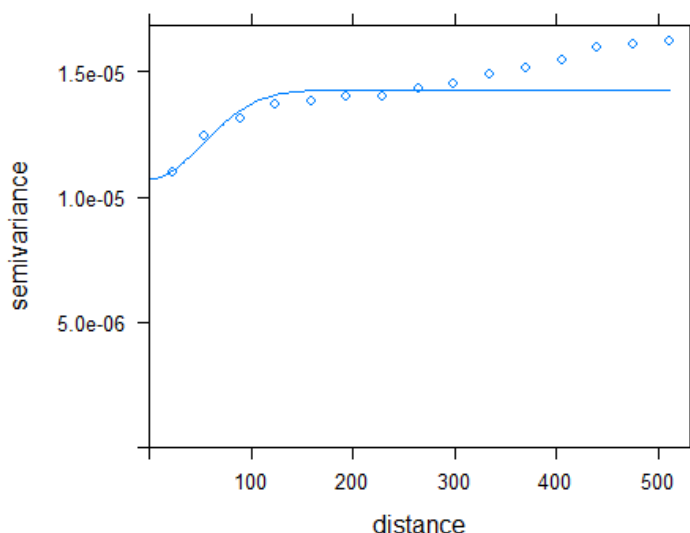


Рис. 4.27. Гаусівська варіограмна модель даних про розподіл миш'яку.

Порівняємо зважену суму квадратичної помилки для кожної моделі. Для експоненційної:

```
attr(fit.variog.exp, "SSErr")
## [1] 3.289497e-11
```

Для сферичної:

```
attr(fit.variog.sph, "SSErr")
## [1] 6.627119e-11
```

Для гаусівської:

```
attr(fit.variog.gau, "SSErr")
## [1] 7.501577e-11
```

Тепер для кожної моделі проведемо прогнозування за допомогою ординарного крігінгу на поверхні України та візуалізуємо результат за допомогою ggplot.

Експоненційна модель(рис. 4.28):

```
krig.exp<-krige(formula=AS1 ~ 1, locations=spdf, newdata=grd,
model=model.variog.exp)
krig.exp.output=as.data.frame(krig.exp)
names(krig.exp.output)[1:3]<-c("long","lat","var1.pred")
plot<-ggplot(data=krig.exp.output,aes(x=long,y=lat))
layer1<-c(geom_tile(data=krig.exp.output,aes(fill=var1.pred)))
layer2<-c(geom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40",
size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint =
median(krig.exp@data$var1.pred,na.rm = T))+labs(fill = "Миш'як", title = "Розподіл
миш'яку по Україні (експоненційна модель)")
```

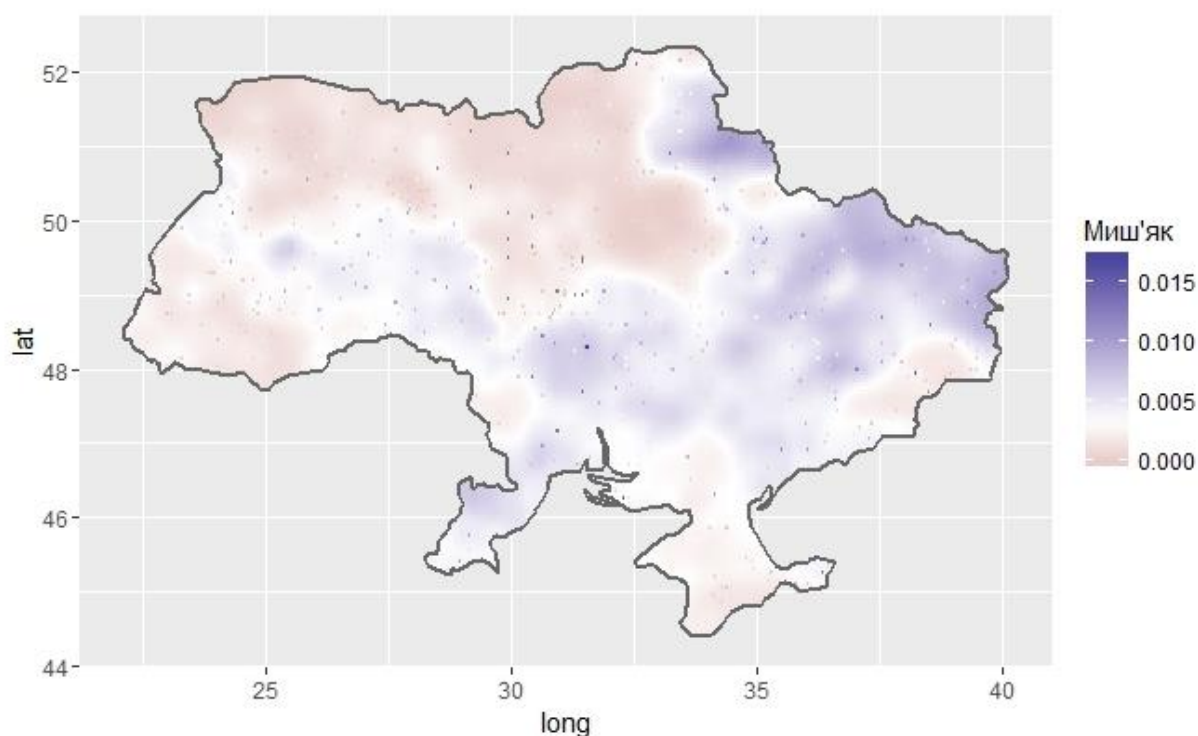


Рис. 4.28. Експоненційна модель розподілу значень миш'яку

Сферична модель (рис. 4.29):

```
krig.sph<-krige(formula=AS1 ~ 1, locations=spdf, newdata=grd, model=model.variog.sph)
```

```
krig.sph.output=as.data.frame(krig.sph)
names(krig.sph.output)[1:3]<-c("long","lat","var1.pred")
plot<-ggplot(data=krig.sph.output,aes(x=long,y=lat))
layer1<-c(geom_tile(data=krig.sph.output,aes(fill=var1.pred)))
layer2<-c(geom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40",
size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint =
median(krig.sph@data$var1.pred,na.rm = T))+labs(fill = "Миш'як", title = "Розподіл
миш'яку по Україні (сферична модель)")
```

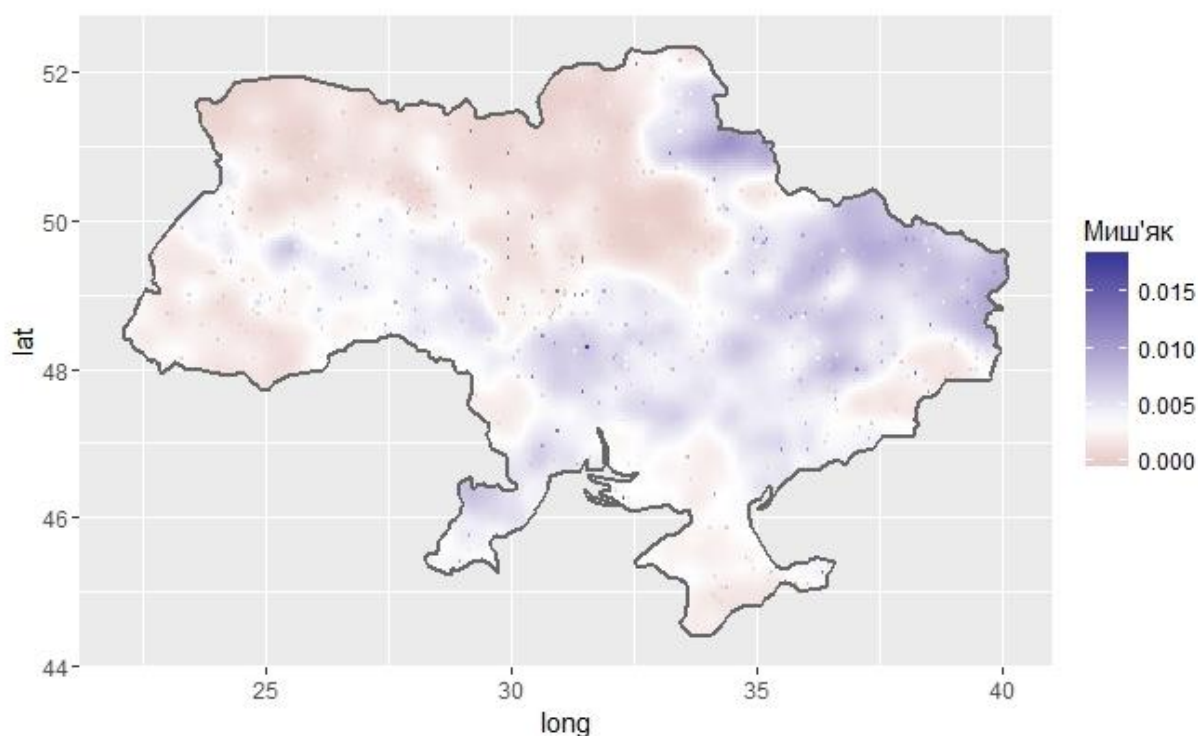


Рис. 4.29. Сферична модель розподілу значень миш'яку

Гаусівська модель (рис. 4.30):

```
krig.gau<-krige(formula=AS1 ~ 1, locations=spdf, newdata=grd, model=model.variog.gau)
```

```
krig.gau.output=as.data.frame(krig.gau)
names(krig.gau.output)[1:3]<-c("long","lat","var1.pred")
plot<-ggplot(data=krig.gau.output,aes(x=long,y=lat))
layer1<-c(geom_tile(data=krig.gau.output,aes(fill=var1.pred)))
layer2<-c(geom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40",
size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint =
median(krig.gau@data$var1.pred,na.rm = T))+labs(fill = "Миш'як", title = "Розподіл
миш'яку по Україні (гаусівська модель)")
```

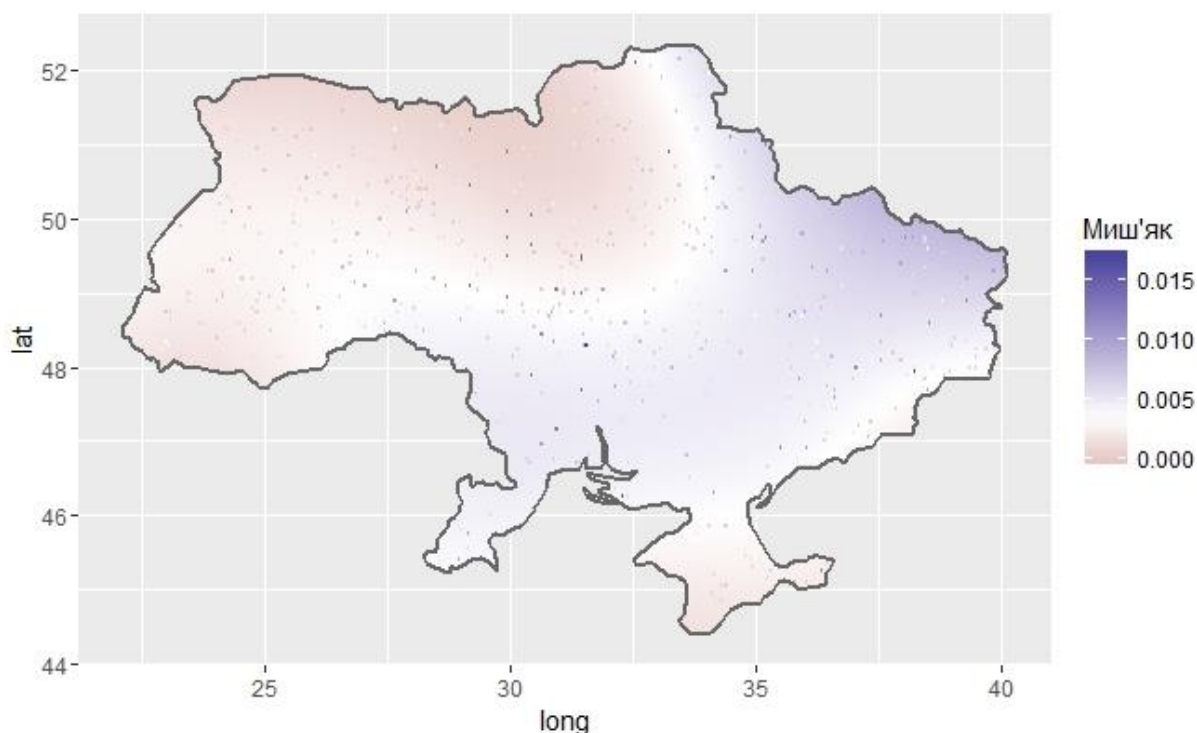


Рис. 4.30. Гаусівська модель розподілу значень миш'яку

Метод крігінга в локальному околі

В класичному крігінгу всі просторові розподіли отримані з урахуванням всіх спостережень для кожного випадку. В більшості випадків зручніше використовувати дані лише в околі точки, для якої здійснюється просторове прогнозування. Причини цього є статистичними та обчислювальні. Статистичні причини включають в себе гіпотезу про те, що стала середня або середня функція повинні застосовуватися локально, або що передбачуване знання варіограми діє тільки до невеликої відстані. Обчислювальні проблеми полягають в швидкості та пам'яті: крігінг при даних розміру n вимагає розв'язку системи $n \times n$. При великих n обчислення виходить повільним, а використовуючи оточуючий окіл можна отримати суттєвий виграш у часі.

Ординарний крігінг з використанням для інтерполяції найближчих сусідів (рис. 4.31):

```
krig.neig10<-krige(formula=AS1~ 1, locations=spdf, newdata=grd, model=model.variog.exp, nmax =10)
krig.neig10.output=as.data.frame(krig.neig10)
names(krig.neig10.output)[1:3]<-c("long", "lat", "var1.pred")
plot<-ggplot(data=krig.neig10.output,aes(x=long,y=lat))
layer1<-c(geom_tile(data=krig.neig10.output,aes(fill=var1.pred)))
layer2<-c(geom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40", size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint = median(krig.neig10@data$var1.pred,na.rm = T))+labs(fill =
"Миш'як", title = "Розподіл миш'яку по Україні (Ординарний крігінг)")
```

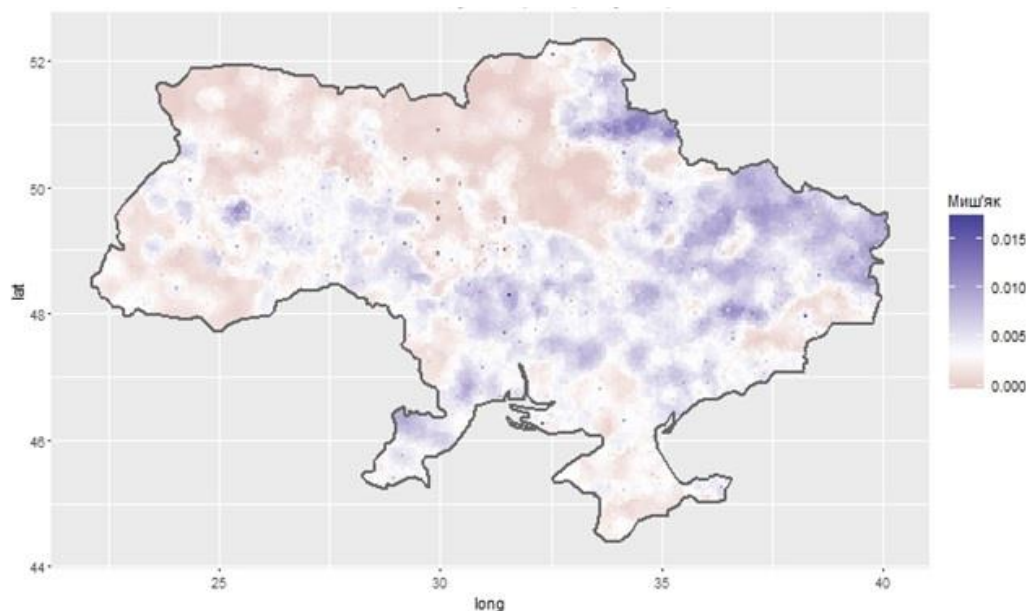


Рис. 4.31. Модель ординарного крігінку для розподілу значень миш'яку
Кокрігінг з використанням для інтерполяції найближчих сусідів (рис. 4.32):

```
krig.neig20<-krige(formula=AS1 ~ 1, locations=spdf, newdata=grd, model=model.variog.exp, nmax =20)
krig.neig20.output=as.data.frame(krig.neig20)
names(krig.neig20.output)[1:3]<-c("long", "lat", "var1.pred")
plot<-ggplot(data=krig.neig20.output,aes(x=long,y=lat))
layer1<-c(ggeom_tile(data=krig.neig20.output,aes(fill=var1.pred)))
layer2<-c(ggeom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40", size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint = median(krig.neig20@data$var1.pred,na.rm = T))+labs(fill =
"Миш'як", title = "Розподіл миш'яку по Україні (кокрігінг)")
```

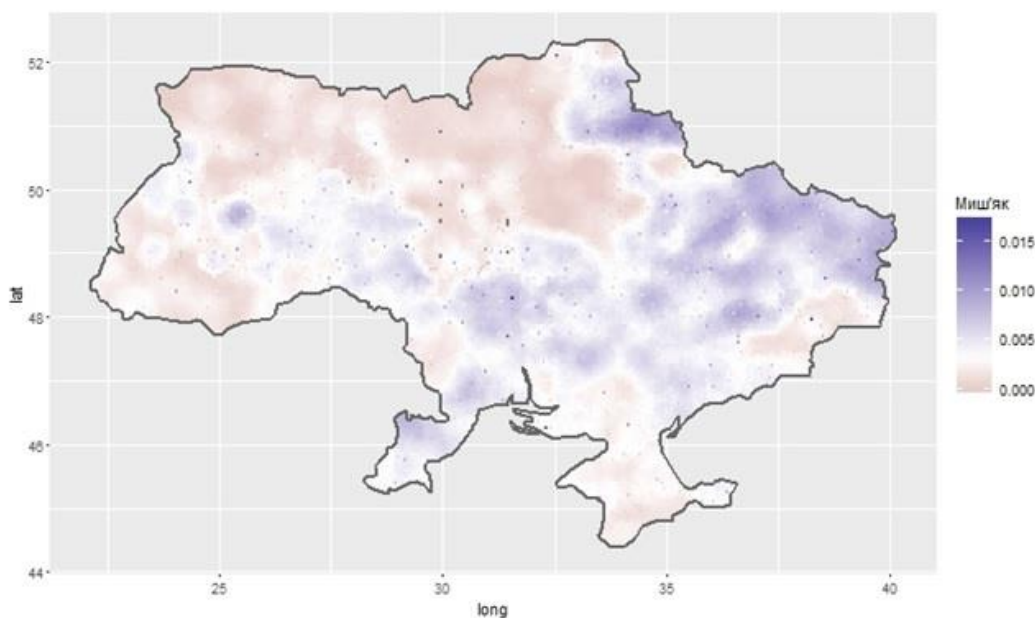


Рис. 4.32. Модель кокрігінгу для розподілу значень миш'яку
Індикаторний крігінг з використанням для інтерполяції найближчих сусідів (рис. 4.33):

```
krig.neig30<-krige(formula=AS1 ~ 1, locations=spdf, newdata=grd, model=model.variog.exp, nmax =30)
krig.neig30.output=as.data.frame(krig.neig30)
names(krig.neig30.output)[1:3]<-c("long", "lat", "var1.pred")
```



```
plot<-ggplot(data=krig.neig30.output,aes(x=long,y=lat))
layer1<-c(geom_tile(data=krig.neig30.output,aes(fill=var1.pred)))
layer2<-c(geom_path(data=Ukraine,aes(long, lat, group=group),colour = "grey40", size=1))
plot+layer1+layer2+scale_fill_gradient2(midpoint = median(krig.neig30@data$var1.pred,na.rm = T))+labs(fill =
"Миш'як", title = "Розподіл миш'яку по Україні (індикаторний крігінг)")
```

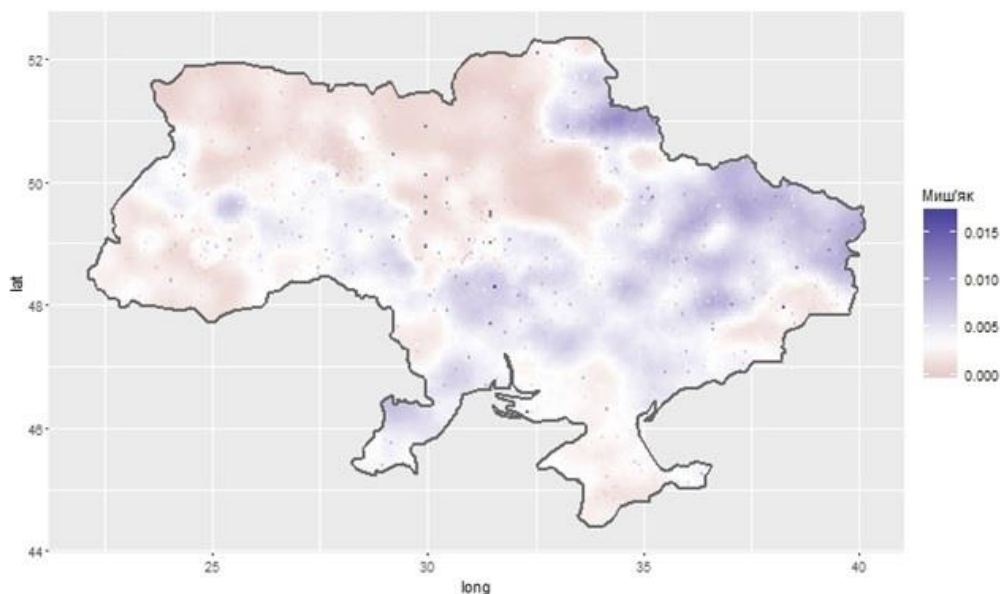


Рис. 4.34. Модель індикаторного крігінгу для розподілу значень миш'яку

Діагностика моделей (Додаток Б)

Таблиця 4.3

Оцінка точності детермінованих та стохастичних інтерполяційних моделей

<i>Метод</i>	<i>Сер. квадр.похибка</i>	<i>Коеф. детермінації</i>
ОЗВ, степінь 1	0.003682343	0.1687547
ОЗВ, степінь 2	0.003577271	0.549687
ОЗВ, степінь 3	0.003934212	0.5831523
Крігінг, експ.	0.003400866	0.4884747
Крігінг, гаус.	0.003592975	0.3921688
Крігінг, сфер.	0.003395039	0.502267
Крігінг, ордин.	0.003462349	0.7788645
Кокрігінг	0.003409049	0.5306521
Індикаторний крігінг	0.003413332	0.5115193

В результаті було з'ясовано, що найменшу квадратичну похибку надає використання методу сферичного крігінгу. В той же час найбільший коефіцієнт детермінації має метод ординарного крігінгу.

Висновки до розділу

Вирішення проблеми сталого розвитку України з врахуванням безпекової складової на основі розробленої методології вимагає розробки методів геостатистичного аналізу просторових даних, що характеризують складові сталого розвитку. Складність такого аналізу обумовлена декількома особливостями: наявністю великих обсягів кількісної та якісної інформації, великим діапазоном масштабів, мінливістю та наявністю різнотемпових факторів впливу.

Розв'язання вказаної проблеми базується на принципі детермінованості у просторі подібності характеристик в залежності від функції розподілу автокореляційних зв'язків між місцем та величиною значення.

Визначено підходи до аналізу та інтерполяції просторово розподілених даних, які умовно розділені на три групи:

- детерміновані моделі (інтерполятори) - лінійна інтерполяція на основі тріангуляції, метод зворотних відстаней в ступені, мульти-квадратичні рівняння;
- стохастичні моделі, що базуються на імовірнісній інтерпретації даних;
- моделі, що засновані на навчанні - штучні нейронні мережі, генетичні алгоритми, статистична теорія навчання машин векторів підтримки.

Розроблено методику ІАГД для розрахунків складових сталого розвитку із застосуванням сукупності методів геостатистичного аналізу, що включає в себе конкуруючі набори структур на основі лінійних і нелінійних методів, параметричних і непараметричних моделей для аналізу, обробки та подання просторової інформації.

Розроблена структуризація та синтез методів геостатистичного аналізу та штучного інтелекту для розв'язання вказаної проблеми.

В методиці розроблено метод застосування трендових поверхонь на основі картометричних обчислень визначення куту нахилу поверхонь та експозиції схилів.

Розв'язана задача представлення просторового тренду. Встановлено наявність, функціональний та кількісний зв'язок вкладу складових індексу в тренд

сталого розвитку для складових якості та безпеки життя людей на території України. Отримано функціональний та кількісний зв'язок просторових розподілів тренду кожного індексу.

Побудовані детерміновані функціональні моделі просторово-часових взаємозв'язків індексу сталого розвитку та його складових та виявлено часові тренди.

Побудовані стохастичні моделі просторово-часових взаємозв'язків. Виконано прогноз імовірності появи величин відповідних показників.

Проведено порівняння між детермінованими та стохастичними моделями в задачі оцінювання концентрації розподілу миш'яку у підземних водах на території України для компонента екологічної безпеки.

Встановлено, що основною геостатистичною моделлю, яка в тому чи іншому вигляді використовується в усіх методах геостатистики, є стохастична модель крігінг (kriging) - лінійний інтерполятор, що використовує для отримання оцінки значення функції в деякій точці простору на основі експериментально виміряних значень цієї функції в інших точках.

Застосування методу ординарного крігінгу та обернено зваженої відстані для моделювання квазіповерхонь розподілу значені вимірів якості та безпеки життя показало переважання методу ординарного крігінгу при моделюванні безперервних розподілів.

На основі даних про вміст миш'яку у підземних водах на території України було виконано порівняльний аналіз детермінованих та стохастичних методів інтерполяції. В результаті було удосконалено застосування методу крігінгу на основі підбору сферичної моделі варіограми, що дозволило отримати найкращі статистичні оцінки для інтерпольованої поверхні. Визначено, що стохастичні методи інтерполяції дають статистично кращі оцінки за показниками середньої квадратичної похибки та коефіцієнту детермінації.

РОЗДІЛ 5. ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ГЕОПРОСТОРОВИХ ДАНИХ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

5.1. Технології штучного інтелекту в ГІС

Однією з основних задач ГІС, яку ставлять перед собою дослідники, є отримання нових знань, уявлень про природу просторових даних. У той же час користувачі іноді недооцінюють можливостей ГІС в області підтримки прийняття рішень, які ці системи можуть забезпечувати, приділяючи основну увагу поданням, зокрема візуалізації даних. Цінність географічної інформації в системах підтримки прийняття рішень стає особливо значущою, коли в ГІС включаються програмні засоби, що базуються на технологіях і методах штучного інтелекту – «розділу інформатики, що вивчає методи, способи і прийоми моделювання і відтворення за допомогою ЕОМ розумної діяльності людини, пов'язаної з вирішенням завдань».

Експертні системи - перший комерційно значущий продукт в області штучного інтелекту. Важливість таких засобів штучного інтелекту як експертні системи і нейронні мережі, полягає в тому, що вони істотно розширюють коло практично важливих задач, які можна вирішувати на комп'ютерах, і їх рішення приносить значний економічний ефект.

Результатом успішного розвитку методів і технологій штучного інтелекту стало створення численних додатків, орієнтованих на кінцевих користувачів, включаючи фахівців в області ГІС. Інтеграція систем штучного інтелекту з ГІС особливо ефективна в задачах оцінки, контролю та прийняття рішення.

В математиці засоби багатовимірного просторового аналізу постійно розвиваються і добре забезпечені методами дослідження геометрії, топології та інших властивостей абстрактних об'єктів, їх множин і структур. Тому їх більш широке використання в геоінформаційних технологіях все більш актуально і здійснюється шляхом:

- 1) розширення функціональної повноти традиційних методів, технології та програмних засобів просторового аналізу в ГІС за рахунок використання можливостей розвинених математичних методів аналізу багатовимірних даних;

2) розвитку нових методів, заснованих на інтелектуальних обчислювальних технологіях, як бази для створення наступного покоління зручних інструментальних програмних засобів аналізу геоданих в умовах великих обсягів первинної інформації;

3) створення нових моделей даних, інформаційних технологій і програмних засобів, спеціально призначених для багатовимірного аналізу даних, моделювання і прогнозу територіально розподілених процесів.

Першим шляхом ГІС розвиваються практично з моменту їх появи. Другий і третій напрямки пов'язані з фундаментальними дослідженнями на стику математики, інформатики та нейрофізіології. За останні роки дослідники добре розробили і виділяють цілий клас статистичних та адаптивних методів аналізу багатовимірних даних, які отримали назву «нейромережеві методи». Нейромережеві методи застосовуються не тільки для аналізу даних, а також для побудови моделей процесів, що розгортаються в багатовимірних просторах. Статистичні та адаптивні методи аналізу геоданих дозволяють поліпшити якість вихідної інформації і побудувати нейромережеву модель, адекватну як призначенням і якістю вихідних даних, так і судженням експертів, і завданням дослідників.

Основи створення нейронних мереж.

Сучасні комп'ютери влаштовані по так званій схемі фон Неймана, що реалізує швидкі послідовності великого числа бінарних операцій. В якості основної альтернативи підходу фон Неймана обговорювалася орієнтація на відтворення принципів роботи біологічних нейронних мереж. Приблизно в ті ж роки створення першого комп'ютера була створена перша нейроподібна система - персептрон Розенблатта.

Порівняємо біологічний нейрон з найбільш часто розглядається схемою технічного нейрона. Обидва типи нейронів реагують на вплив з боку багатьох нейронів в залежності від величини зв'язків з цими нейронами. На відміну від технічних нейронів реакція біологічного нейрона завжди невід'ємна, причому, якщо вплив на нього не досягнув критичного рівня, реакції немає. Можливо, з цим

пов'язана одна з найбільш очевидних відмінностей біологічних нейронних мереж від існуючих сьогодні нейропрограм - один і той же мозок в залежності від того, які нейрони «мовчать», може працювати досить різним чином.

Типи технічних нейромереж.

У літературі помітна увага приділяється питанням архітектури технічних нейронних мереж, наведемо варіант відповідної класифікації схем: одношарові, багатошарові і повнозв'язні.

Не менш важливим є розбиття нейроалгоритмів на два класи. У першому випадку навчання організовано як відтворення набору правильних зразків (навчальної вибірки), після чого мережа може адекватно реагувати і на приклади, яких не було в навчальній вибірці, у другому випадку зразки правильної реакції початково відсутні.

Нейромережі, які навчаються за зразком, походять від персептронів і в сучасному трактуванні можуть розглядатися як варіанти і модифікації мереж зі зворотним поширенням помилки (іноді як результат примітивізації такого роду мереж з метою спрощення реалізації). До цього класу можна віднести, наприклад, одношаровий і багатошаровий перцептрон, машину Больцмана, мережі, які навчаються за правилом Хебба, рекурентні шаруваті і повнозв'язні мережі зворотного поширення помилки, мережі, що використовують радіальні базисні функції.

Нейромережі, які навчаються без зразку різноманітніші, хоча теорія стоїть за ними математично часом більш примітивна - це карти Кохонена, системи з множинними локально стійкими станами, такі, як мережа Хопфілда, мережі, що налаштовуються на основі адаптивного резонансу.

Списки завдань, які прийнято відносити до найчастіше розглянутих в рамках нейромережевого підходу:

- навчання за зразком - класифікація образів, апроксимація функцій, передбачення, управління, аналіз даних, категоризація всередині класу, стиснення даних;

- навчання без зразку - категоризація, категоризація всередині класу, аналіз даних, стиснення даних, асоціативна пам'ять.

Географічний підхід до вивчення явищ природи і суспільного життя передбачає територіальну мінливість цих явищ і її вивчення за допомогою методів класифікації. Районування території, типологія і оцінка комплексів найчастіше з поданням на мапі отриманих результатів є не тільки методами, але і цілями досліджень. Більшість аналітичних задач такого роду можна сформулювати як задачу класифікації географічних комплексів. При цьому попередньо здійснюється класифікація показників і чинників, що описують ці комплекси. У цих дослідженнях нейромережеві алгоритми можуть бути використані для відновлення функції за кінцевим набору значень і для розбиття кінцевої множини об'єктів на класи.

Для змістовної інтерпретації потрібен аналіз близькості виділених класів між собою і виявлення їх зв'язку з різними сторонами досліджуваного явища. Для цього використовуються різні алгоритми факторного або кореляційного аналізу. Використання цих методів неефективно при великих обсягах, а особливо при нелінійному характері взаємозв'язків в первинних даних. У цьому випадку доцільно використання нейромережевих алгоритмів, які можна інтерпретувати як узагальнення методів лінійної статистики на нелінійний або локально-лінійний випадок.

Дуже важливим підставою географічних класифікацій є тимчасові, генетичні (спадкові) характеристики досліджуваних комплексів. У географічних дослідженнях комплекси зазвичай розглядатися як просторово-часові освіти. Тимчасові і генетичні характеристики відіграють важливу роль і в класифікаціях, що зустрічаються в інших природних і суспільних науках (біології, геології, економіці, історії та ін.) Результати яких відображаються на географічних картах. Використовувана при цьому багатовимірна класифікація та критерії достовірності алгоритмів аналізу багатовимірних даних можуть бути незастосовні до реальних географічним завданням і поданням відповідних даних в ГІС. Щодо гнучкі нейромережеві підходи можуть виявитися більш ефективними.

Однією з істотних проблем аналізу геоданих є вибір оптимальної системи базових показників відповідно до суті досліджуваних явищ. Природне бажання всебічного розгляду явища в максимально повній системі показників може призвести до надмірності включених в аналіз первинних даних. Дані, що поставляються з різних джерел і подаються як незалежні, можуть багаторазово дублюватися або обчислюватися одні на основі інших (похідні дані). Це може спотворити значимість ознак і привести до помилок в результатах аналізу. Найсуттєвішу допомогу тут може надати глибоке проникнення в сутність досліджуваного комплексу територіально-розподілених процесів, що дозволяє знайти критерій значущості показників і сформувати їх систему адекватну дослідницької задачі. В автоматизованому режимі можливе використання методів он-лайн корекції моделей і техніки обчислювального експерименту, що забезпечують виправлення помилок в даних, автоматизоване виявлення причинно-наслідкових зв'язків, зниження розмірності задач багатовимірної класифікації та аналізу геоданих.

Показники лежать в основі більшості класифікаційних задач аналітичної географії мають різну природу і можуть носити як кількісний так якісний характер. Тому алгоритми аналізу геоданих повинні вміти працювати з характеристиками як числовий так і нечислової природи. Це накладає певні обмеження на можливості використання всього різноманіття методів математичного аналізу. У ГІС атрибутивний опис об'єктів аналізу вже формалізовано і представлено у формі електронних таблиць. Це полегшує використання нейромережевих технологій і дозволяє подолати труднощі аналізу, пов'язані з великими обсягами первинної інформації, пробілами в даних, різноманітністю кількісних і якісних характеристик об'єктів. Нейромережевий підхід дозволяє проводити компактизацію і параметризацію даних, створювати на основі даних прості і наочні моделі даних.

Перелік наведених проблем аналізу територіально-розподілених процесів, представлених у вигляді географічних комплексів фактично виводить на необхідність використання методів нейромережевого аналізу для структуризації багатовимірних даних. Виявлення об'єктів або їх комплексів (вони самі по собі теж

об'єкти зі своєю складною структурою), встановлення між ними різних просторових відносин і заходів близькості. Комплексування об'єктів та його узагальнення даних істотно залежить від постановки дослідницького завдання і дозволяє значно скоротити число об'єктів аналізу. При переході від аналітичних задач до задач оцінки і прогнозу стає визначальним побудова моделей досліджуваних процесів. При цьому:

- виявлення в результаті аналізу об'єкти стають базовими елементами моделі,
- показники географічного комплексу переходять в числові показники властивостей об'єкта і його зв'язків,
- аналіз географічних комплексів і їх показників виводить на визначення функціональних зв'язків і можливості структурування об'єктів відповідно до їх функціональними зв'язками.

Для побудови моделей важливий також аналіз причинно-наслідкових відносин подій (зміна станів об'єктів залучених в той чи інший процес), структура функціональних зв'язків виділених об'єктів, показники обмінних процесів реалізуються на цих зв'язках, визначення характеристик польового типу і т.д.

Вельми істотно і те, що часто трудомісткість рішення задачі традиційними методами і на основі нейроалгоритмів суттєво різняться. У ситуації, коли необхідно вникнути в структуру даних, потім написати програму обробки даних, потім налагодити цю програму, буває досить просто завантажити «сирі» дані в нейропрограму, почекати деякий час (якщо база даних велика, наприклад, почекати ніч), і на ранок отримати готові результати.

Одним з ефективних засобів рішення слабо формалізованих задач на основі прикладів є нейронні мережі. Поширення нейронних мереж пояснюється наступними перевагами нейросетевого підходу:

1. Автоматична настройка параметрів нейросетевой моделі для вирішення завдання на прикладах. Не потрібно участь експерта для побудови моделі, вирішальною завдання.

2. Універсальність. Нейросети дозволяють стандартним чином, без урахування семантики, вирішувати будь-які завдання, які допускають представлення у вигляді набору прикладів, що містять вхідні і вихідні дані.

3. Стійкість при роботі з зашумленими і недостовірними даними.

4. Можливість адаптації (донавчання) до нових умов.

5. Стійкість до збоїв і руйнувань елементів.

6. Високий паралелізм, властивий нейромережевим моделям.

7. Здатність ефективно обробляти дані високої розмірності, різнотипні дані.

Потрібно також відзначити, що нейромережі в сукупності з геоінформаційними технологіями можуть використовуватися для вирішення завдань, що виходять за рамки традиційних додатків ГІС. Наприклад, вони знаходять застосування при візуалізації і картографуванні багатовимірних даних (в яких можуть міститися пробіли) за допомогою вкладених в простір даних двовимірних многовидів. Особливістю технології є можливість безперервного проектування даних на карту, що істотно підвищує точність представлення даних. Основою для побудови такої карти є двовимірна прямокутна сітка, вложена в багатовимірний простір, апроксимуюча дані і володіє регульованими властивостями пружності по відношенню до розтягування і вигину - тому її називають пружною картою. Розташування вузлів сітки шукається в результаті рішення варіаційної задачі на знаходження мінімуму функціоналу, вид і параметри якого залежать від дослідницької мети і визначають пружні властивості карти.

На відміну від «традиційних» статистичних методів, нейромережі видають не статистично достовірне, а правдоподібне рішення задачі і можуть застосовуватися при нестачі емпіричних даних для статистичного дослідження. В якості переваг нейронних мереж в порівнянні зі статистичним підходом можна назвати універсальність і автоматизований режим настройки в умовах сильної апріорної невизначеності, що дозволяє швидко отримати прийнятний результат. Нейромережеві моделі накладають слабкі обмеження на можливі функції розподілу змінних і дозволяють уникати апріорних припущень про вид функцій розподілу змінних і структурі моделі.

У літературі помітна увага приділяється питанням архітектури технічних нейронних мереж, наведемо варіант відповідної класифікації схем (рис. 5.1):

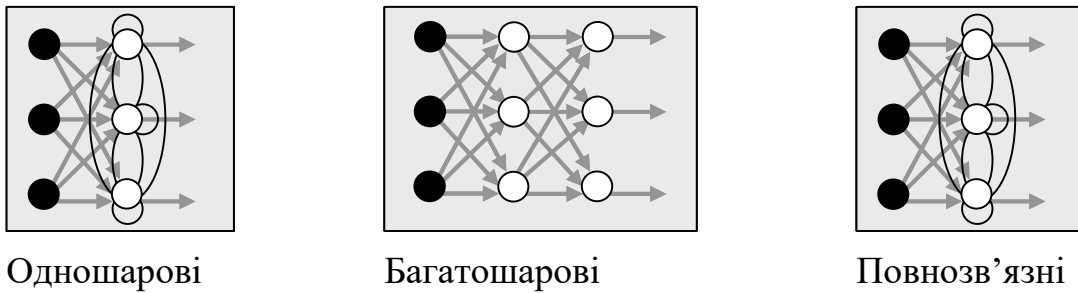


Рис. 5.1. Класифікація типів нейронних мереж

Не менш важливим є розбиття нейроалгоритмів на два класи - Supervised (навчаються за зразком, з Учителем) і Unsupervised (навчаються без зразка, без Вчителі). У першому випадку навчання організовано як відтворення набору правильних зразків (навчальної вибірки), після чого мережа може адекватно реагувати і на приклади, яких не було в навчальній вибірці, у другому випадку зразки правильної реакції початково відсутні.

Нейроалгоритми зворотного поширення помилки початково створювалися багато в чому для вирішення класичної завдання математичної статистики - завдання регресії табличних даних. Добре відомі найпростіші завдання регресії - такі як завдання проведення прямої, найкращим чином наближає хмара точок (наприклад, прямий, сума квадратів відстаней точок до якої мінімальна). У цьому прикладі мова йде про лінійної регресії, алгоритм якої повинен визначити параметри положення прямої - її кути нахилу до осей координат і координати однієї з точок прямої.

Якщо мова йде про нелінійну регресію, яка наближає хмару точок різноманіття виступає не пряма, що не площину або гіперплощина, а деяка гладка крива, поверхня або гіперповерхня. Щоб описати такі нелінійні різноманіття, потрібно збільшити число параметрів у порівнянні з випадком лінійної регресії: виникає завдання багатопараметричної нелінійної регресії [76].

Математично все виглядає наступним чином: є неявна функція $H(\beta)$, наприклад задається співвідношеннями

$$H = H(x(i)), \psi_j(x_i, \beta_k) = 0, k = 1, \dots; N_{ij} = 1, \dots M; (1)$$

тобто є M величин x_i стільки ж рівнянь на них, причому в ці рівняння входять параметри β , а функція H явно залежить від x . Якщо ми навчимося вирішувати загальну задачу про знаходження мінімуму функції $H(\beta)$, то і завдання багатопараметричної нелінійної регресії вирішимо, оскільки, як ми побачимо згодом на прикладах, це завдання як правило може бути представлена у формі

Введемо множники Лагранжа μ_j і похідну функцію

$$W(\alpha, \beta, \mu) = H(\alpha) + \sum_{j=1}^M \mu_j \cdot \psi_j(\alpha, \beta) \quad (5.1)$$

За допомогою рівняння на x , що входять в, можна записати у формі

$$\partial W / \partial \mu_j = 0, j = 1, \dots, M \quad (5.2)$$

Якщо (як передбачатиметься нижче), умови виконані, то $W=H$, отже

$$\begin{aligned} dH / d\beta_i &= dW / d\beta_i = \partial W / \partial \beta_i + \\ &+ \sum_{j=1}^M (\partial W / \partial \alpha_j \cdot \partial \alpha_j / \partial \beta_i + \partial W / \partial \mu_j \cdot \partial \mu_j / \partial \beta). \end{aligned} \quad (5.3)$$

Щоб знайти екстремуми функції $H(\beta)$, у багатьох алгоритмах потрібно мати ефективний спосіб обчислення повних похідних $H(\beta)$ за параметрами β . вираз стало результатом тотожних перетворень. Якщо максимально спростити цей вираз, його можна буде використовувати для знаходження $dH/d\beta_i$. Зауважимо, що останні члени можна не враховувати - вони рівні нулю. До сих пір множники Лагранжа μ були довільними введеними чисто формально параметрами. За рахунок цього, вибравши в якості додаткових умов на множники Лагранжа рівняння

$$\partial W / \partial \alpha_j = 0, j = 1, \dots, M, \quad (5.4)$$

можна перетворити на нуль і передостанні члени в (5.4), звівши (5.4) до простішого вигляду

$$dH / d\beta_i = \partial W / \partial \beta_i \quad (5.5)$$

і дають шуканий простий спосіб обчислення dNd/β_i . Якщо зіставляти наведені співвідношення з простою аналогією, з якої ми починали, то можна сказати що з'явилися додаткові рівняння описують «потік в штучно перевернутої системі», якщо вважати що описують «потік у вихідній системі» [112].

Спеціаліст з прикладної математики дізнається в наведених міркуваннях стандартну схему, ще від Лагранжу, і широко використовувану у багатьох додатках - наприклад, в теорії оптимального управління. Природно, що використання такого підходу в задачах неявно заданої багатопараметричної регресії виявилось ефективним.

5.2. Застосування нейронних мереж без вчителя для аналізу геопросторових даних

З метою підтримки досліджень глобальних змін і розвитку системних моделей функціонування Землі, Міністерство науки і технологій Китаю запустило Національну високотехнологічну програма досліджень та розробок «Високі технології досліджень і розробки Програми Китаю» (863 Program) під назвою «Картографування за допомогою даних ДЗЗ і дослідження ключових технологій побудови глобального рослинного покриву» у 2010 році. Одним з результатів проекту є картографічний продукт GlobeLand30-2010 глобального покриву з 30-метровий просторовим розрізненням, отриманий із зображень дистанційного зондування в 2010 році [326].

Набір даних охоплює територію площею від 80° Пн. ш. до 80 ° Пд. ш. та складається з 10 типів земного покриву, а саме сільськогосподарські землі, лісові, степові, деревно-чагарникові, водно-болотні угіддя, водні об'єкти, тундра, урбаністичні ландшафти, пустельні ландшафти та льодовики [140].

Для генерації даних земного покриття використовувалися класифіковані зображення із супутникових даних мультиспектральних зображень GlobeLand30-2010 з роздільною здатністю 30 м, в тому числі Landsat TM і ETM+ і мультиспектральні зображення китайської системи Environmental Disaster

Alleviation Satellite (HJ-1). Були обрані безхмарні зображення, отримані протягом вегетаційного сезону в межах ± 1 рік відносно 2010 року [140].

Усі зображення Landsat TM і ETM + були завантажені з ресурсу Геологічної служби США (USGS, <http://landsat.usgs.gov/>) на рівні 1Т. До знімків Landsat були застосовані геометричні і радіометричні виправлення зображень, які потім були зареєстровані в системі координат WGS-84 системи з проекцією UTM. Таким чином, були завантажені 9907 зображення Landsat TM і ETM + зображень з USGS, близько 80% з них були придбані в період з 2009 по 2011 р [140].

Попередня обробка Landsat TM і ETM + зображень включає в себе атмосферну і топографічну корекцію. Атмосферна корекція використовує оновлену версія алгоритму Швидкої лінії візування атмосферного аналізу спектрального Гіперкубу (FLAASH), в той час як топографічна корекція використовує метод згладжених COS (SCOS).

HJ-1 китайський супутник дистанційного зондування, спрямований на моніторинг навколишнього середовища та запобігання і зменшення стихійних лих. HJ-1A супутник несе мультиспектральну камеру (CCD) і гіперспектральний сенсор (HSI). Супутник HJ-1B несе мультиспектральну камеру (CCD) та інфрачервоний мультиспектральний сенсор (IRS) (<http://www.cresda.com/>) [140].

1465 сцен мультиспектральних зображень HJ-1 було використано для отримання інформації про земний покрив, покриваючи 60% від загальної площі земної кулі, з часом придбання, починаючи з вересня 2008 року по грудень 2011 року (рис. 5.2) [140].

Мінімальна одиниця відображення (іменована надалі «МОВ») відноситься до найменшої ділянки, яка використовується для контролю якості. Площа ділянки більше розміру МОВ повинна пройти через процес контролю якості. Що стосується лінійних об'єктів, таких як річки, для контролю застосовується мінімальна ширина. МОВ варіюється від типу до типу земного покриття, який визначається наступними факторами:

- ◆ Просторовий розподіл об'єктів (розмір, щільність і т.д.);
- ◆ Масштабні характеристики типу об'єктів;

♦ Обсяг роботи людино-машинного інтерактивного редагування та перевірки;

♦ Мінімальні вимоги до точності класифікації об'єктів.

Продукт GlobeLand30-2010 було отримано з використанням ієрархічного методу вилучення. Кожен тип рослинного покриву був класифікований по одному, з обмеженнями маскою інших типів рослинного покриву. Робочий процес полягає в тому щоб отримати тільки один тип земного покриву за один раз, а потім маскувати клас після детекції. Потім класифікація здійснюється для наступного типу рослинного покриву, який потім маскується, поки не будуть отримані всі класи [140].

Загальна модель обробки даних представлена на рис. 5.2.



Рис. 5.2. Алгоритм аналізу адміністративних одиниць за показниками землекористування з використанням нейронних мереж.

Порядок отримання типів земного покриву є наступним (рис. 5.2):

- 1) Водойми;
- 2) Водно-болотні угіддя;

- 3) Урбанізовані території;
- 4) Сільськогосподарські землі;
- 5) Постійні сніги та лід;
- 6) Ліси, чагарники і луки;
- 7) Тундра

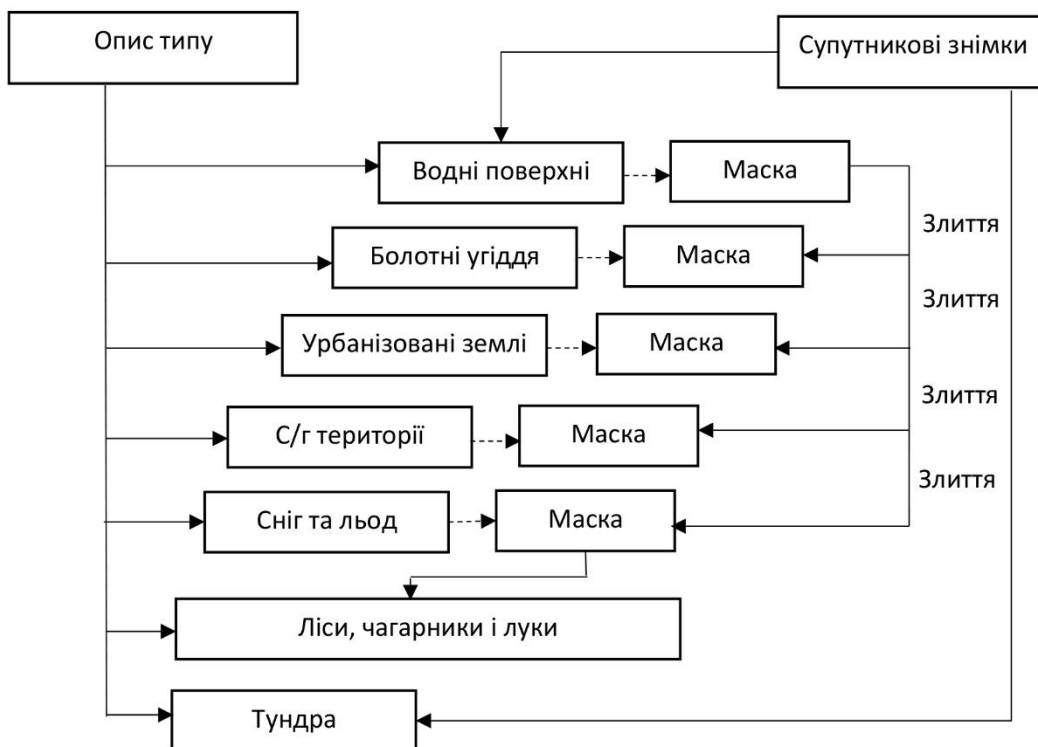


Рис.5.3. Ієрархічна стратегія класифікація типів

Відповідно до карти тайлів та карт кожного континенту було обрано 8 тайлів для оцінки точності. Загалом, 9 типів земного покриття та 150000 тестових прикладів було оцінено з точки зору точності. Загальна точність GlobeLand30-2010 досягає 83.51%. Індикатор Каппа дорівнює 0,78 [140].

GlobeLand30-2010 набір даних має наступні характеристики системи координат.

- ◆ система координат: WGS84
- ◆ Референц-еліпсоїд: WGS84
- ◆ Проекція: UTM
- ◆ Зонування: 6°

GlobeLand30-2010 продукт організований в тайлах:

- ◆ 60 ° Пн. - 60 ° Пд.: 5° (широта) x 6° (довгота);
- ◆ 60-80° Пн. і 60-80° Пд: 5° (широта) x 12° (довгота).

Кількість тайлів карти GlobeLand30-2010 є 853.

Тайлові дані GlobeLand30-2010 складаються з п'яти частин, а саме файлу результату класифікації, файлу інформації про координати, індексного картографічного файлу класифікації зображень, файлу метаданих та ілюстративного файлу. Серед них:

- Файл класифікації результату: відповідає файлу зберігання інформації про класифікацію земного покриву.
- Файл інформації про координати відповідає файлу запису координатної інформації результатів класифікації даних.
- Індексний картографічний файл класифікації відповідає файлу векторного шару запису діапазону і часу кожної сцени класифікаційних зображень.
- Файл метаданих відповідає файлу запису інформації про метадані результатів класифікації;
- Ілюстративний файл відповідає файлу, що описує результат класифікації даних і пояснення авторських прав на дані [140].

Система класифікації включає в себе десять видів земного покриву, а саме сільськогосподарські землі, ліси, луки, чагарник, водно-болотних угіддя, водойми, тундра, урбанізовані території, пустелі, льодовий покрив. Будь ласка, зверніться до таблиці 1 для ілюстрації кожного типу:

Таблиця 5.1

Класифікація GlobeLand30

Код	Тип	Зміст
10	Сільськогосподарські землі	Землі, які використовуються для сільського господарства, багаторічних насаджень, в тому числі рисових полів, зрошуваних сільгоспугідь, овочівництва, фруктових садів.
20	Ліс	Землі, які вкриті лісовою рослинністю з покривом більше 30%, в тому числі листяні і хвойні ліси, і рідколісся з покриттям 10 - 30% і т.д.

30	Луки	Землі, які покриті натуральною трав'янистою рослинністю з покриттям більше 10%, і т.д.
40	Деревно-чагарникова рослинність	Землі, які вкриті чагарниками з покриттям більше ніж 30%, в тому числі листяні і вічнозелені чагарники і пустельний степ з покриттям більше 10%, і т.д.
50	Водно-болотні угіддя	Землі, які вкриті водно-болотною рослинністю і водоймами, в тому числі внутрішні болота, заболочені озера, заплави річок з водно-болотними угіддями, ліси / чагарники водно-болотних угідь, торф'яних болота і т.д.
60	Водні об'єкти	Водні об'єкти на суші, в тому числі річки, озера, водосховища, ставки для розведення риби і т.д.
70	Тундра *	Землі вкриті лишайником, мохом, витривалою багаторічною трав'янистою рослинністю і чагарниками в полярних регіонах, в тому числі чагарникової тундри, трав'янистих тундрах і т.д.
80	Урбанізовані території	Землі модифіковані діяльністю людини, в тому числі всі види житла, промислові і гірничодобувні галузі, транспортні комунікації, внутрішні міські зелені зони і водойми і т.д.
90	Пустельні території *	Земля з рослинним покривом менше 10%, в тому числі пустелі, піщані поля, Гобі, голі скелі і т.д.
100	Постійний сніговий або льодовий покрив*	Земля вкрита вічним снігом, льодовики і крижаний покрив.

* - відсутність даного типу земного покриття в Україні або наявність у кількості менше 1 % [140]

Підготовка даних для території України складалась з наступних етапів:

- Завантаження тайлів, що покривають територію України
- Обрізка країв тайлів за допомогою векторних масок

- Злиття мозаїки
- Обрізка мозаїки по контуру державного кордону України
- Експорт даних у формат GeoTIFF.

На наступному етапі необхідно сформувати геостатистичні дані для адміністративних одиниць України на різних територіальних рівнях. Перевагою такого підходу є можливість формування статистичних показників для різних територіальних одиниць або їх поєднань: річкові басейни, природні зони та ін [160].

Формування геостатистичних даних відбувається з використанням операцій зональної статистики, які реалізовано в модулі Spatial Analyst ПЗ ArcGIS. Для обробки даних якісного розподілу класифікаційних ознак підходить інструмент «Зональна статистика в таблицю», який дозволяє резюмувати значення растру у межах зон іншого набору даних [140].

Зональні інструменти дозволяють виконувати аналіз, вихідні дані якого є результатом обчислень, виконаних на всіх комірках, що належать кожній вхідній зоні. Зона може бути визначена як область однакових значення, але вона також може складатися з декількох відокремлених елементів, або регіонів, всі з яких мають одне значення. Зони можуть задаватися растром або наборами класів векторних об'єктів. Растр повинен мати тип ціле число, а просторові об'єкти повинні мати цілочисельне або строкове поле атрибутів [140].

Деякі зональні інструменти оцінюють кількість певних властивостей геометрії, або форми, вхідних даних зони, і їм не потрібні інші вхідні дані. Інші зональні інструменти використовують вхідні дані зони для визначення місць розташування, для яких будуть обчислюватися інші параметри, наприклад, статистика, площі або частота значень. Існує також зональний інструмент, який використовується для заповнення зазначених зон мінімальними значенням, що знаходяться на межі зони [140].

Таблиця 5.2

Інструменти для роботи з зонами діляться на дві категорії з точки зору застосування [140]

Категорія інструментів для роботи з зонами	Інструменти
Інструменти, що працюють з формами зони.	Зональна геометрія (Zonal Geometry) Зональна геометрія в таблицю (Zonal Geometry As Table)
Інструменти, що працюють з атрибутами зони.	Зональна статистика (Zonal Statistics) Зональна статистика в таблицю (Zonal Statistics as Table)
Інструменти, що визначають частку класів в межах зони.	Таблиця площ (Tabulate Area)
Інструменти, що визначають розподіл частот значень одного вхідного растру в зонах, визначених іншим растром.	Зональна гістограма (Zonal Histogram)
Інструменти, що заповнюють зазначені зони.	Заповнення зон (Zonal Fill)

Отримані значення площ можна використовувати для оцінки розподілу фактичних площ землекористування, а також для дослідження співвідношення між різними типами землекористування у відсотках, що потребує нормалізації даних відносно загальної площі адміністративної одиниці [140].

Математичною основою моделі є штучна нейронна мережа на основі карт самоорганізації Кохонена (рис. 5.4). Карта самоорганізації, зазвичай, представляється як двовимірна матриця масивів вузлів нейронів. У нашому випадку даний масив формує гексагональну решітку, що є ефективним рішенням для візуального аналізу даних [140].

Процедуру роботи карт самоорганізації описано в роботах Т. Кохонена, С. Хайкіна, Ф. Уоссермена та інших.

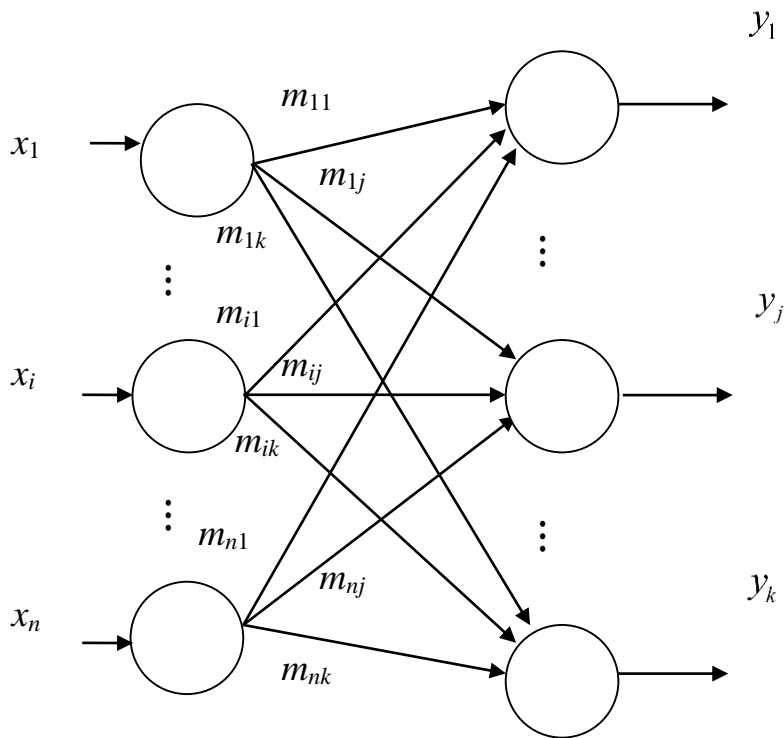


Рисунок 5.4. Шар нейронів Кохонена

Опишемо процедуру побудови даної нейронної мережі. Вхідний простір даних $X \subset \mathbb{R}^n$ проектується, найчастіше, на двовимірну решітку вузлів нейронів. Кожному вузлу i , $i = \overline{1, k}$, ставиться у відповідність параметричний вектор моделі $m_i = \{\mu_{1i}, \mu_{2i}, \dots, \mu_{ni}\} \in \mathbb{R}^n$, де k - кількість вузлів нейронів. У той же час вхідний вектор $x = \{x_1, x_2, \dots, x_n\} \in X \subset \mathbb{R}^n$ пов'язаний одночасно з усіма нейронами через скалярні ваги μ_{ji} [140].

Перед початком проведення навчання нейронної мережі всі ваги μ_{ji} ініціалізуються за допомогою генератора випадкових величин, але слід зазначити, що існують й інші методи [140].

Після ініціалізації мережі запускається так званий процес конкуренції, який полягає у визначенні найбільш подібних нейронів до векторів вхідних даних. Найбільш поширеним підходом до даного процесу є обчислення Евклідової відстані:

$$\|x - m_i\| = \sqrt{\sum_{j=1}^n (x_j - \mu_{ji})^2}, \quad i = \overline{1, k}, \quad j = \overline{1, n}. \quad (5.6)$$

Після подання вхідних векторів на входи нейронів, відбувається змагання нейронів шару Кохонена за правилом «переможець отримує все», згідно якого обирається нейрон-переможець, що є найближчим до вхідного вектора [140]:

$$c = \arg \min_{i=1,k} \{\|x - m_i\|\}, \quad (5.7)$$

Після визначення нейрона-переможця здійснюється визначення околу його впливу на сусідні нейрони. Нейрон-переможець знаходиться в центрі топологічного околу. При збудженні, він впливає на просторово близькі до нього нейрони, проте даний вплив зменшується із збільшенням відстані між нейронами. Даний процес призводить до визначення топологічного сусідства нейронів [160].

На заключному етапі побудови карти самоорганізації відбувається процес синаптичної адаптації. Даний процес заключається у процедурі корегування вектора синаптичних ваг нейрона-переможця та його сусідів відповідно до вектора вхідного показника:

$$m_i(t+1) = m_i(t) + v(t)h_{ci}(t)[x(t) - m_i(t)], \quad (5.8)$$

де $v(t)$ – параметр швидкості навчання; $h_{ci}(t)$ – функція топологічного сусідства з центром у нейроні-переможці c . Результат налаштування такого типу нейронних мереж можна представити у вигляді карти Кохонена (рис.5.5) [160].

Опишемо властивості такої нейронної мережі. Даний підхід є унікальним – поєднує в собі процедури проектування, кластеризації та виявлення взаємозв'язків між даними. Головною особливістю такого типу нейронних мереж є навчання «без вчителя». Це означає, що для коректного налаштування синаптичних ваг не потрібний бажаний відгук мережі, а отже процес моделювання протікає без втручання суб'єктивної думки розробника [160].

Оскільки карти самоорганізації здійснюють стиснення інформації із збереженням найбільш важливих топологічних та/або метричних зв'язків між вхідними елементами даних, можна також вважати, що з їх допомогою створюються абстракції (узагальнення) певного виду. Ці дві властивості карт самоорганізації, візуалізацію та узагальнення, можна використовувати у вирішенні таких складних задач, як аналіз, управління та передача інформації [160].

В якості вхідних даних використовували розподіл земель між класами землекористування для кожної статистичної одиниці. В якості таких одиниць були обрані адміністративні райони України. Таким чином на вхід було подано 8 параметрів, які характеризують земний покрив України за виключенням пустельних районів та постійного снігового покриву. Відповідно до кількості вхідних параметрів було обрано кількість вихідних кластерів, припускаючи, що кожний додатковий клас вносить додаткове збурення та може виступати центром нового кластеру. В результаті було отримано 8 розподілів для кожного параметру та загальну кластеризацію на основі карт самоорганізації Кахонена. (рис. 5.5) [160].

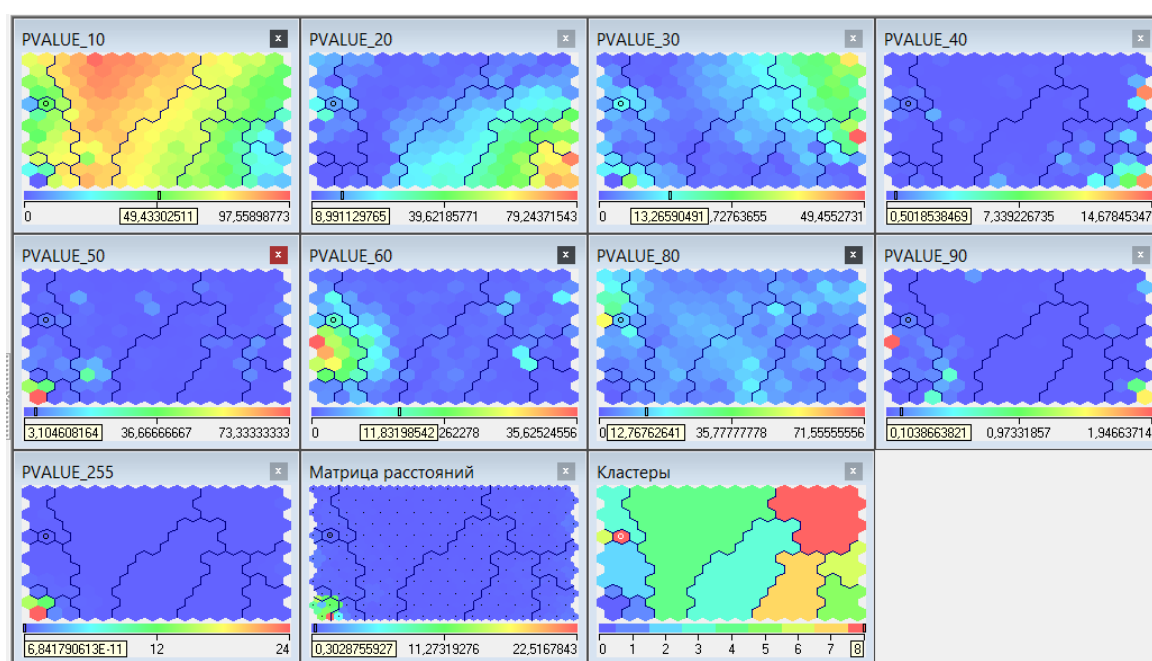


Рисунок 5.5. Карти самоорганізації Кахонена на основі даних про земний покрив

Проведення кластеризації дозволило виділити кластери у географічному просторі шляхом розфарбування адміністративних районів за кольорами якісного фону на основі належності до відповідного кластеру. Як видно з рисунку 3 значна кількість районів має сусідів, що відносяться до того ж кластеру, що показує зональність у розподілі кластерних ознак. Це, звичайно, закономірно, оскільки в основі аналізу покладено співвідношення між різними типами земного покриву, що в свою чергу залежить від таких зональних факторів у розподілі як кліматичні ознаки та рослинний покрив. Азональними факторами в цьому випадку виступають

антропогенно змінені території, які додають мозаїчність у просторовий розподіл кластерів (рис.5.6) [140].

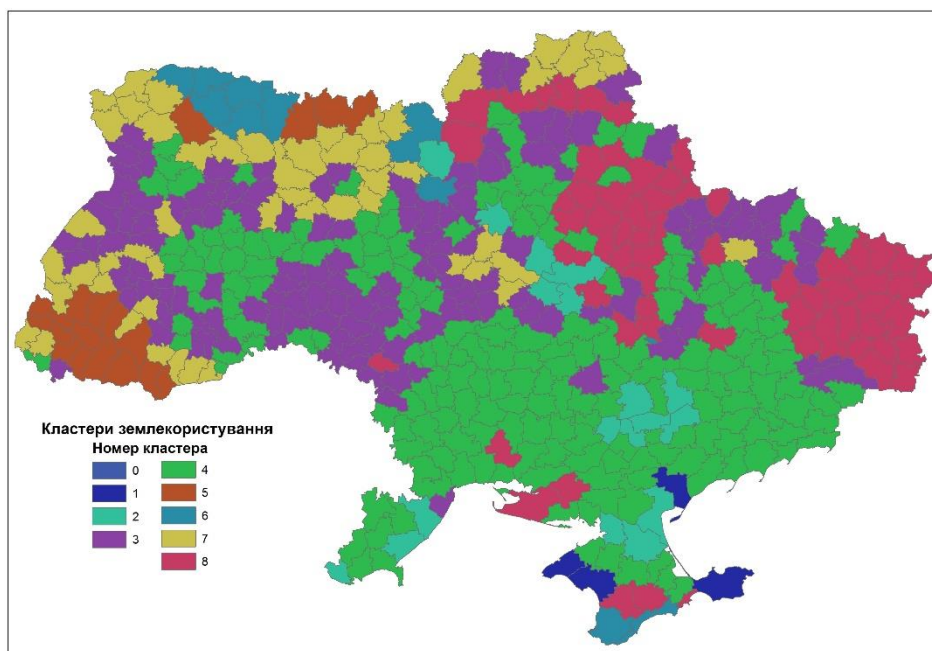


Рисунок 5.6 Кластеризація адміністративних районів на основі даних про земний покрив

Кластер 0 характеризується вологими природними ландшафтами з переважанням лісових та болотних угідь.

Кластер 1 характеризується пустинними степовими ландшафтами півдня країни зі значним відсотком антропогенного втручання.

Кластер 2 визначається наявністю значних водних об'єктів у межах адміністративних одиниць, які у більшості випадків відносяться до водосховищ Дніпровського каскаду. Таким чином, визначним видом функціонування даних районів є водне господарство.

Кластер 3 характерний для лісостепової зони і характеризується рівномірним співвідношенням між такими групами земного покриву як ліси, сільськогосподарські землі, урбанізовані території, лучна рослинність. Спеціалізація кластерів: сільськогосподарське виробництво та переробка, машинобудування, рекреація та охорона природи.

Кластер 4 зосереджений у своїй більшості на півдні України та на Поділлі. Кластер характеризується домінуванням у структурі землекористування сільськогосподарських земель, зокрема рілля та сіножатей, а також сільських населених пунктів. Кластер має компактну структуру, яка залежить від сприятливих агрокліматичних умов.

Кластер 5 відноситься до зон поширення хвойних лісів та чагарників. Зосередження кластеру спостерігається у Карпатських горах та у Поліссі. Основні види діяльності пов'язані з лісовим господарством та екстенсивними формами тваринництва. Також активно розвивається рекреаційна діяльність та охорона природи.

Кластер 6 відноситься до кластерів що характеризуються високим різноманіттям форм землекористування з переважанням урбанізованих територій та лісових масивів.

Кластер 7 має у своєму складі переважання мішаних та листяних лісів, чагарників, луків з невеликим відсотком антропогенно змінених територій. Основні види діяльності пов'язані з лісовим господарством, тваринництвом, рекреацією.

Кластер 8 характеризується переважанням сільськогосподарських земель та урбанізованих територій з високим відсотком змінених ландшафтів. Основними функціями цих кластерів є промислове виробництво та переробка сільськогосподарської продукції [160].

5.3. Параметризація складових сталого розвитку з використанням показників нічної освітленості в Україні

У 2015 році Україна приєдналась до глобальних зусиль у досягненні Цілей сталого розвитку (ЦСР) в рамках Саміту ООН для прийняття Порядку денного в галузі розвитку на період після 2015 року. Цей інтегральний план дій побудовано на основі гармонізації економічного, екологічного та соціально-інституційного вимірів розвитку задля, зокрема, подолання злиденності в усіх її формах й

посилення боротьби зі зміною клімату. Глобальна система оцінки ЦСР охоплює близько 230 показників [365], отримання високоякісних даних за якими часто стає викликом для національних статистичних органів.

Адекватні дані безумовно мають важливе значення для прийняття обґрунтованих рішень та відслідковування прогресу на шляху досягнення ЦСР. Це вимагає збирання, обробки, аналізу і розповсюдження безпрецедентного обсягу даних та статистичної інформації на усіх рівнях, зокрема і даних з нових та інноваційних джерел, як вказано у звіті “ЦСР 2016” [387]. У цьому ж звіті особливо наголошено на важливості дослідження нових джерел даних, нових технологій збирання даних, інтеграції різних джерел даних, зокрема, поєднання геопросторової інформації та статистичних даних. У глобальних звітах зі сталого розвитку [257, 336] неодноразово наголошувався потенціал застосування так званого підходу “великих даних” (англ. “big data”) у вимірюванні прогресу зі сталого розвитку, тобто використання супутникового дистанційного зондування і комунікаційних технологій.

Яскравим прикладом імплементації такого підходу може слугувати програма ARSET (скор. від Applied Remote Sensing Training) НАСА [385], в рамках якої проходять тренінги з упровадження супутникових даних в прийняття рішень. Наприклад, один з вебінарів ARSET 2017 року було присвячено використанню супутникових даних з вмісту завислих часток розміром менше 2,5 мкм для оцінювання якості повітря з огляду на показники ЦСР 3 “Підтримання здоров’я” та ЦСР 11 “Сталі міста” [368], ще один - з ґрунтово-рослинного покриву для аналізу прогресу на шляху досягнення ЦСР 15 “Забезпечення життя на землі”.

Визначення економічного стану країни або регіону — питання досить непросте. Один зі способів непрямого оцінювання економічної активності ґрунтується на аналізі зміни освітленості на досліджуваній території в нічний час на основі супутникових знімків [200]. У роботі [275] оцінено економіку африканських країн на південь від Сахари. У роботі [230] зроблено спробу оцінити економічну ситуацію на Сході України за допомогою вивчення зміни освітлення за період військових дій, а в роботі [18] — на основі даних про споживання

електроенергії на цих територіях, отриманих із супутникових знімків. Інтенсивність освітленості тісно пов'язана з важливим показником економічної активності — споживанням електроенергії. Автори відповідної методики Дж. Хендерсон з колегами [275] встановили, що в країнах з низьким і середнім рівнем доходу на душу населення змінення нічної освітленості на 1 % приблизно дорівнює зміненню доходів на 1 %.

У нашій роботі ми використовували дані зйомок із супутника Suomi NPP, який було виведено на орбіту в 2011 р.

Супутник має 5 бортових сенсорів:

- 1) Advanced Technology Microwave Sounder (ATMS);
- 2) Cross-track Infrared Sounder (CrIS);
- 3) Ozone Mapping and Profiler Suite (OMPS);
- 4) Visible Infrared Imaging Radiometer Suite (VIIRS);
- 5) Clouds and the Earth's Radiant Energy System (CERES).

Для тематичної обробки даних зйомок було використано продукт Day/Night Band (DNB), розроблений на основі знімків з приладу VIIRS, — місячний композит нічної освітленості території з нічних знімків [412].

За допомогою програми ArcGIS 10.4 було здійснено статистичний аналіз аналіз освітленості території з метою порівняння з інтегральними індексами економічного виміру, який застосовується в оцінці сталого розвитку.

Формування геостатистичних даних відбувається з використанням операцій зональної статистики, які реалізовано в модулі Spatial Analyst ПЗ ArcGis. Для обробки даних якісного розподілу класифікаційних ознак підходить інструмент «Зональна статистика в таблицю», який дозволяє резюмувати значення растру у межах зон іншого набору даних [412].

Зональні інструменти дозволяють виконувати аналіз, вихідні дані якого є результатом обчислень, виконаних на всіх комірках, що належать кожній вхідній зоні. Зона може бути визначена як область однакових значення, але вона також може складатися з декількох відокремлених елементів, або регіонів, всі з яких мають одне значення. Зони можуть задаватися растром або наборами класів

векторних об'єктів. Растр повинен мати тип ціле число, а просторові об'єкти повинні мати цілочисельне або строкове поле атрибутів [412].

Деякі зональні інструменти оцінюють кількість певних властивостей геометрії, або форми, вхідних даних зони, і їм не потрібні інші вхідні дані. Інші зональні інструменти використовують вхідні дані зони для визначення місць розташування, для яких будуть обчислюватися інші параметри, наприклад, статистика, площі або частота значень. Існує також зональний інструмент, який використовується для заповнення зазначених зон мінімальними значенням, що знаходяться на межі зони [412].

Таблиця 5.3.

Інструменти для роботи з зонами діляться на дві категорії з точки зору застосування [412]

Категорія інструментів для роботи з зонами	Інструменти
Інструменти, що працюють з формами зони.	Зональна геометрія (Zonal Geometry) Зональна геометрія в таблицю (Zonal Geometry As Table)
Інструменти, що працюють з атрибутами зони.	Зональна статистика (Zonal Statistics) Зональна статистика в таблицю (Zonal Statistics as Table)
Інструменти, що визначають частку класів в межах зони.	Таблиця площ (Tabulate Area)
Інструменти, що визначають розподіл частот значень одного вхідного растру в зонах, визначених іншим растром.	Зональна гістограма (Zonal Histogram)
Інструменти, що заповнюють зазначені зони.	Заповнення зон (Zonal Fill)

В якості базової одиниці зональної статистики було обрано адміністративно-територіальний поділ України за областями. В результаті були отримані

статистичні дані освітленості нічними вогнями для кожного регіону. Ці дані були порівняні з даними моделі оцінювання сталого розвитку. В результаті було отримано матрицю кореляційних зв'язків між індикаторами та показниками загроз та середнім значенням освітленості для регіону.

Методологія оцінювання та аналізу сталого розвитку, яка викладена в роботах [4] і використовується в цьому дослідженні, включає модель сталого розвитку, яка є міждисциплінарним узагальненням моделей, відомих в природничих, економічних та соціальних галузях науки, та методику застосування формальних статистичних методів та методів експертного оцінювання для аналізу процесів сталого розвитку (Розділ 1.1.).

З метою визначення сили зв'язності між показниками сталого розвитку та освітленості регіонів було відібрано ключові показники, що характеризують індекс сталого розвитку.

В результаті було отримано кореляційну матрицю перевірки зв'язності між статистичними показниками освітленості регіонів та їх відображенням в основних параметрах сталого розвитку.

Таблиця 5.4.

Значення кореляції для показників освітленості та складових індексу сталого розвитку за регіонами України [355]

Значення освітленості			
Складові індексу сталого розвитку	Max	Min	Mean
Компонента якості життя	-0.3062362	0.6958459	0.775869
Компонента безпеки життя	-0.0870352	0.209811661	0.5144
Економічне благополуччя	-0.3388719	0.031903556	0.053557
Індекс екологічного виміру	-0.4851694	0.030830184	-0.12258
Індекс економічного виміру	0.16050311	0.66635917	0.837093
Індекс людського розвитку	-0.0076972	0.086519007	-0.0017
Індекс конкуренто спроможності	0.19815784	0.573736492	0.766881
Індекс якості життя	-0.2961539	0.113519109	0.140841
Індекс соціально-інституціонального виміру	-0.3372712	0.29152067	0.317593
Індекс сталого розвитку	-0.2552884	0.590410857	0.757561

Екологічна безпека	-0.3308463	0.044158404	0.168294
Економічна безпека	0.12583802	0.345941902	0.637313
Безпека (інше)	0.33190022	0.061210228	-0.09038
Вплив на змінювання клімату	-0.0620561	0.77810648	0.997755
Екологічне навантаження і небезпеки	-0.4012843	0.093752585	-0.31661
Стан довкілля	-0.3183306	0.149893923	0.144667

Аналіз цієї матриці дозволяє встановити тісноту зв'язку між освітленістю регіонів та складовими оцінки сталого розвитку. Перевірка статистичних викидів максимальних та мінімальних значень показує відсутність тісних зв'язків. Розрахунки середніх значень, навпаки, вказують на наявність значних залежностей між показниками. Серед економічних показників найбільшу тісноту зв'язку вказує індекс економічного виміру 0.837093, що свідчить про вірність математичної моделі розрахунку. Але найбільший індекс кореляції має показник впливу на змінювання клімату, який дорівнює майже 1 (0.997755). Загроза «Вплив на змінювання клімату» оцінено як неагативний внесок в змінювання клімату кожного регіону України за допомогою показника щільності викидів двоокису карбону:

$$x_{CSH} = \frac{x_{CSH1}}{x_{ACH1}}, \quad (5.9)$$

де x_{CSH} – індекс територіального регіонального внеску в змінювання клімату, т CO₂- екв./км²; x_{CSH1} – викиди двоокису карбону в атмосферне повітря, тис. т. Така висока кореляція між показниками вказує на пряму залежність між викидами CO₂ та освітленням регіонів.

Таблиця 5.5.

Показники освітленості, індексу економічного виміру та загрози впливу на змінюваність клімату.

Регіон	Середнє	Індекс економічного виміру	Вплив на змінювання клімату
Херсонська	0.406302773	0.189978446	0.013081021
Кіровоградська	0.426873212	0.182681374	0.0283342
Житомирська	0.440202611	0.191489871	0.016997697
Чернігівська	0.447090956	0.188034625	0.0221949
Черкаська	0.455947647	0.199617861	0.083043975
Івано-Франківська	0.459405081	0.191448318	0.617942765

Тернопільська	0.494408037	0.154816347	0.0612497
Луганська	0.503027457	0.136353224	0.03580706
Рівненська	0.519161804	0.173368204	0.050337949
Чернівецька	0.521207501	0.128533194	0.099717423
Вінницька	0.522742945	0.256386621	0.106268496
Сумська	0.527845862	0.282721238	0.035942405
Миколаївська	0.534049536	0.221789673	0.044164491
Хмельницька	0.548284084	0.220013636	0.07281069
Одеська	0.56976335	0.31331144	0.044636864
Закарпатська	0.577737223	0.173117245	0.057384757
Полтавська	0.586251418	0.260953541	0.061490813
Запорізька	0.60418225	0.301628786	0.191257325
Волинська	0.621275881	0.152815647	0.028406967
Харківська	0.632036996	0.41468461	0.111485193
Львівська	0.644094097	0.264057942	0.105127533
Дніпропетровська	0.749803804	0.471879995	0.35024505
Київська	0.765244306	0.317667177	0.097460688
Донецька	1.095395932	0.327619324	0.617946673
м.Київ	10.58320843	0.833868434	114.2473272

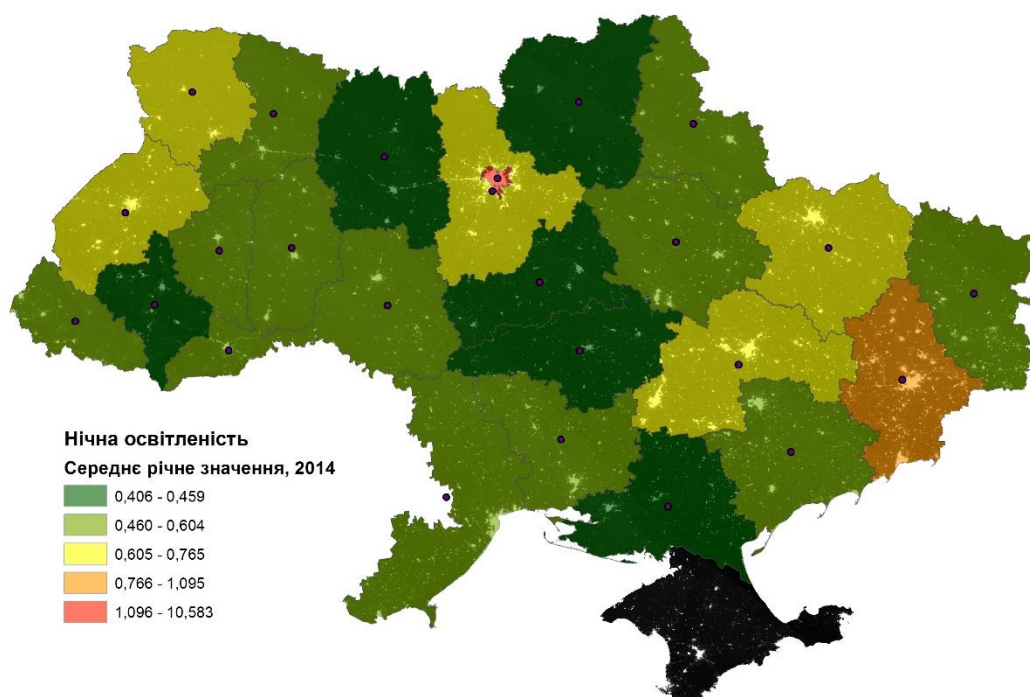


Рисунок 5.7. Середнє значення освітлення окремих регіонів України в 2014 р.

Середній розподіл показників освітлення у 2014 році за регіонами України. Найвище значення має м. Київ. Незважаючи на військові дії Донецька область продовжувала мати високе значення освітлення. Наступні позиції посідають Харківська, Дніпропетровська, Київська, Львівська та Волинська області.

Найменші значення освітлення мають Херсонська, Кіровоградська та Житомирська області.

Для встановлення зв'язку використовуються показники просторової кореляції та просторової регресії.

Географічна зважена локальна регресія визначається як:

$$u_i(x_k, y_k) = \beta_{0i}(x_k, y_k) + \beta_{1i}(x_k, y_k) * v_{1i} + \beta_{2i}(x_k, y_k) * v_{2i} + \dots \\ \dots + \beta_{pi}(x_k, y_k) * v_{pi} + \varepsilon_i, \quad (5.10)$$

для спостереження i , де $u_i(x_k, y_k)$ - залежна змінна, що оцінюється в місці i ; v_{pi} - пояснююча перемінна; $\beta_p(x_k, y_k)$ - локальні коефіцієнти регресії; p кількість перемінних, та ε_i залишок оцінений в місці розташування i . Кожне локальне регресійне $u_i(x_k, y_k)$ рівняння вирішується із різними вагами спостережень, що базуються на функції розпаду з відстанню, що центрована на спостереження i .

Географічно зважений локальний коефіцієнт кореляції Пірсона дозволяє встановити тісноту зв'язку між просторово розподіленими даними. Для цього він розраховується для кожної пари значень двох явищ у місцях їх розміщення за формулою:

$$r_i(x_k, y_k) = \frac{\sum_{j=1}^n \varpi_j * (u_j - \bar{u}) * (v_j - \bar{v})}{(n-1) * S_u * S_v} = \beta_1(x_k, y_k) \frac{S_u}{S_v}, \quad (5.11)$$

де, $r_i(x_{ki}, y_k)$ - це коефіцієнт кореляції, (x_k, y_k) - це місце спостереження i ; u_i та v_i - це окремі спостереження; \bar{u} та \bar{v} - це середні значення двох перемінних, ϖ_j це вага, яку має кожний вимір в залежності від функції розпаду з відстанню, що центрована на спостереження i ; n - це розмір вибірки; S_u та S_v стандартні відхилення двох перемінних; $\beta_1(x_k, y_k)$ - розрахунковий параметр для двовимірної місцевої регресії спостереження i [412].

Для використання методу географічно зваженої регресії та кореляції використовуються перетворення атрибутивних даних з тематичних растрів пожежної небезпеки та щільності розподілу згарищ до точкового векторного шару, що містить атрибутивні ознаки кожного растрового шару. Організація даних у

такій моделі дозволяє застосувати алгоритми географічно-зваженої кореляції на мові Python [355].

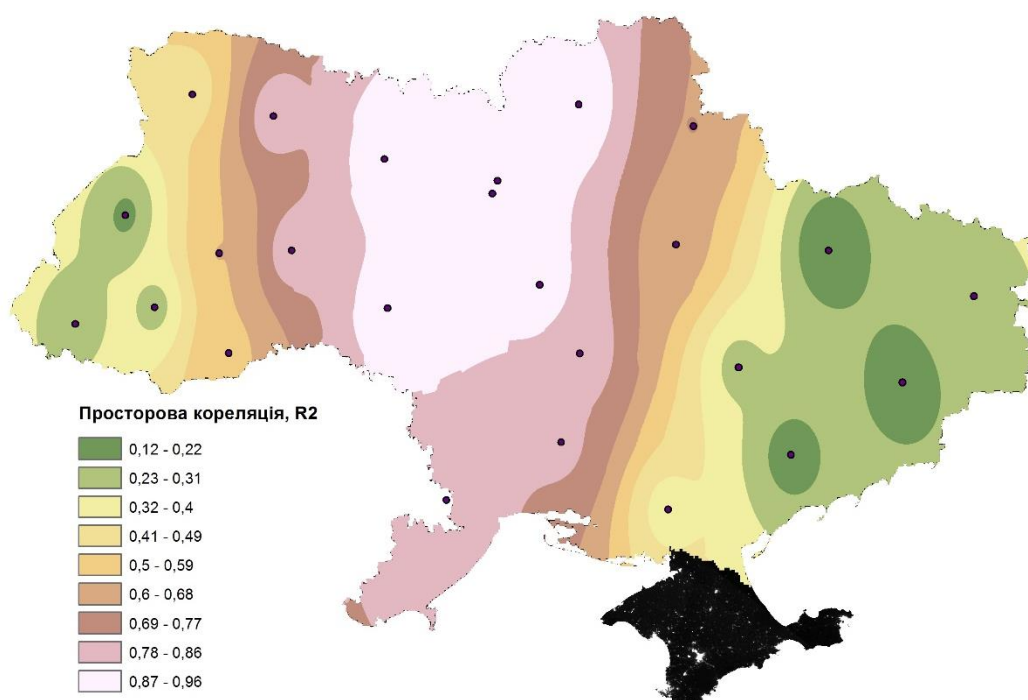


Рисунок 5.8. Значення просторової кореляції регіонів України за показниками середньої освітленості та індексу економічного виміру.

Оцінка просторової кореляції між даними показує наявність найбільш тісного зв'язку між показниками для центральної частини України та її відхилення для східної та західної частини країни. На сході зниження тісноти зв'язку може бути пов'язано з військовими діями та окупацією території. На заході з неврахуванням частини економічних доходів засобами офіційної статистики.

5.4. Застосування нейронної мережі для завдань регіонального аналізу індексу економічного розвитку в моделі сталого розвитку

Трансформаційні процеси в економіці України, що зумовлені плином глобальної економічної кризи, поставили перед керівництвом держави низку завдань, що стосуються, насамперед, підвищення рівня ефективності управління економічними процесами. Методологія сталого розвитку містить інструментарій, який вже апробований, проте постійно удосконалюється. Саме процес

удосконалення зумовлений новими потребами науковців і користувачів, до того ж з'являються нові проблеми саме економічного характеру, які слід враховувати.

Економічний вимір є складовою компоненти якості життя індексу сталого розвитку. Об'єктивна необхідність регулювання суспільного розвитку саме пов'язана з функціонуванням економічних законів. Безпосередньо рівень економічного розвитку є визначником рівня якості життя. Саме при звертанні уваги на цю складову для керівництва регіонального та національного рівнів надається можливість інтегрально оцінити економіку того чи іншого регіону та, при потребі, прийняти ті чи інші управлінські рішення.

Зазначимо, що при аналізі в комплексі з екологічним та соціальним вимірами для осіб, що приймають рішення, надається можливість розглядати проблеми адміністративно-територіальних одиниць з позицій та на засадах концепції сталого розвитку. Відповідно, особа, що приймає управлінські рішення, може визначити попереджувальні заходи, запровадити стабілізаційні дії, започаткувати соціально-економічні реформи на основі лише тих показників, які є доцільним на поточний період.

Розроблені та апробовані методики вже надають можливість здійснювати процес регулювання. Основою методики обрахунку індексу економічного виміру сталого розвитку, яка була запропонована колективом науковців у минулих монографіях [5, 185], були наступні концептуальні засади: 1) всі категорії політики мали однакову вагу; 2) значення вагового коефіцієнта для кожного параметра визначалося експертами, зважаючи, в основному, на його частку у ВВП країни; 3) переважна більшість параметрів була приведена на одну особу (здійснювалися нормування).

Автори методики Дж. Хендерсон з колегами [275] встановили, що в країнах з низьким і середнім рівнем доходу на душу населення змінення нічної освітленості на 1 % приблизно дорівнює зміненню доходів на 1 %.

На основі цієї концепції запропоновано реалізацію розрахунку індексу економічного розвитку на базі даних дистанційного зондування Землі, які найбільше характеризують стан економічного розвитку. Оскільки яскравість

об'єктів на нічних знімках є тільки одним з індикаторів соціально-економічного розвитку, то для більш ґрунтовного аналізу необхідно застосовувати додаткові дані дистанційного зондування.

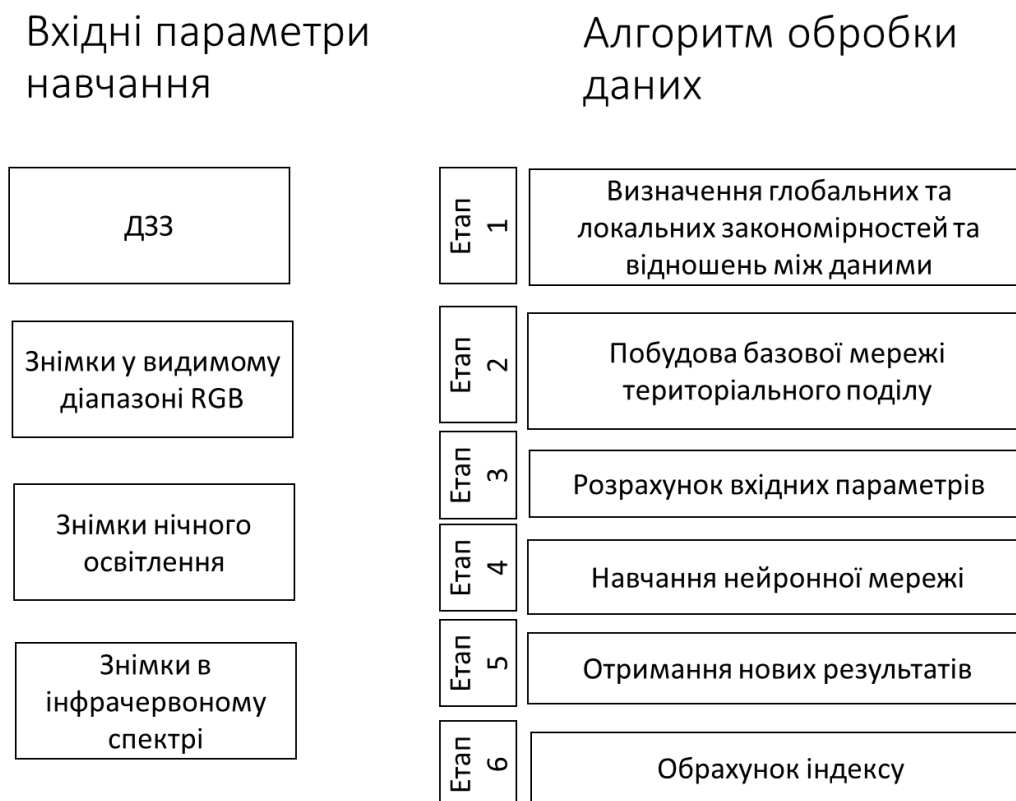


Рис. 5.9. Алгоритм розрахунку індексу економічного розвитку на основі обробки даних ДЗЗ з використанням нейронної мережі

До основних типів даних ДЗЗ, які можуть бути залучені наступні основні категорії даних:

- Знімки середньої роздільної здатності;
- Знімки середньої роздільної здатності нічного часу;
- Знімки високої роздільної здатності;
- Радарні знімки;

Застосування такого підходу дозволяє перейти від макрорівня дослідження сталого розвитку до мікрорівня та отримувати оцінки розвитку території в розрізі районів, об'єднаних територіальних громад, окремих населених пунктів.

Дослідження Стенфордської дослідницької групи продемонстрували, що релевантними ступеню економічного розвитку є знімки середньої та високої

роздільної здатності, які дозволяють ідентифікувати стан інфраструктури та економічного розвитку території та прокорелювати його з індексом економічного розвитку [313].

Згідно з цими рекомендаціями в якості вхідних даних було обрано:

- Знімки нічних вогнів;
- Знімки теплового сліду;
- Знімки у видимому діапазоні для аналізу структури сільськогосподарських земель;
- Знімки у видимому діапазоні для аналізу стану індустріалізації об'єктів.

При цьому можливо використати два підходи до аналізу супутникових знімків:

Класифікація пікселів знімку за визначеним варіантом класифікації.

Детекція об'єктів на космічному знімку за відповідними сигнатурами пікселів.

В даному випадку обрано перший варіант класифікації знімків для подальшого використання.

Для встановлення зв'язків між даними було обрано використання багатошарової нейронної мережі, яка була розроблена у вигляді бібліотеки для програмного середовища ArcGIS 10.6.

Навчання нейронної мережі було реалізовано на основі базових параметрів, які розраховуються на основі космічних знімків.

В якості базових параметрів було обрано:

- середня яскравість нічного освітлення S
- яскравість теплового сліду T
- відсоток рілля у використанні сільськогосподарських земель Z
- відсоток урбанізованих індустріальних територій U

Таким чином залежність між індексом економічного розвитку та геопросторовими параметрами з космічних знімків представлена як

$$f = (S, T, Z, U),$$

$$S, T, Z, U \in \begin{cases} N_l \\ N_m \\ N_h \end{cases},$$

де N_l – космічні знімки низької роздільної здатності

N_m – космічні знімки середньої роздільної здатності

N_h – космічні знімки високої роздільної здатності

Для співставлення космічних даних було обрано знімки середньої роздільної здатності серії Landsat 8 з середнім розрізненням 30 м. та Sentinel 2 з середнім розрізненням 10 м.



a)



б)

Рис. 5.10. Фрагменти знімків у видимому діапазоні а) Landsat 8; б) Sentinel 2A

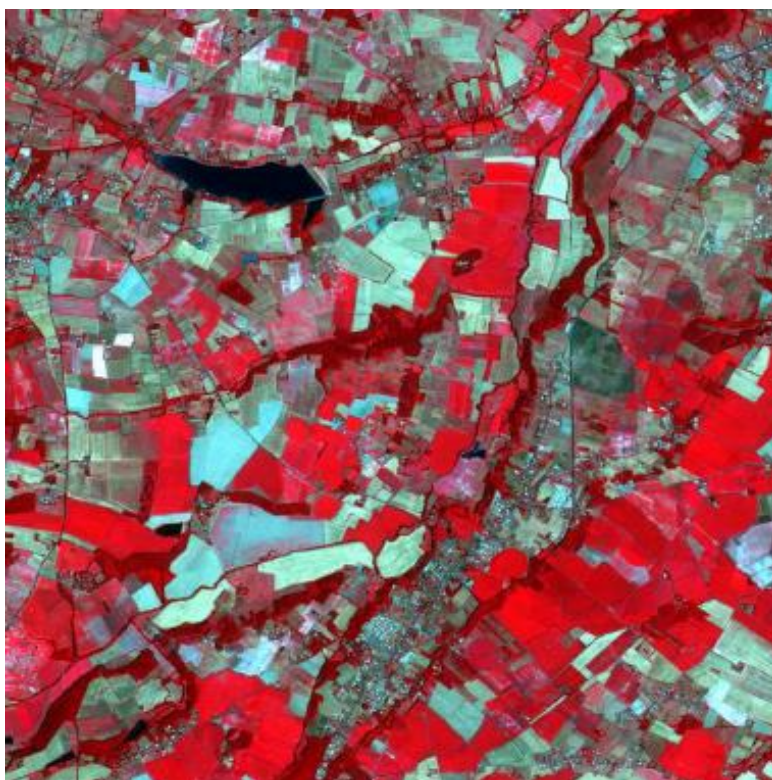


Рис.5.11. Інфрачервоних канал космічного знімку Landsat 8

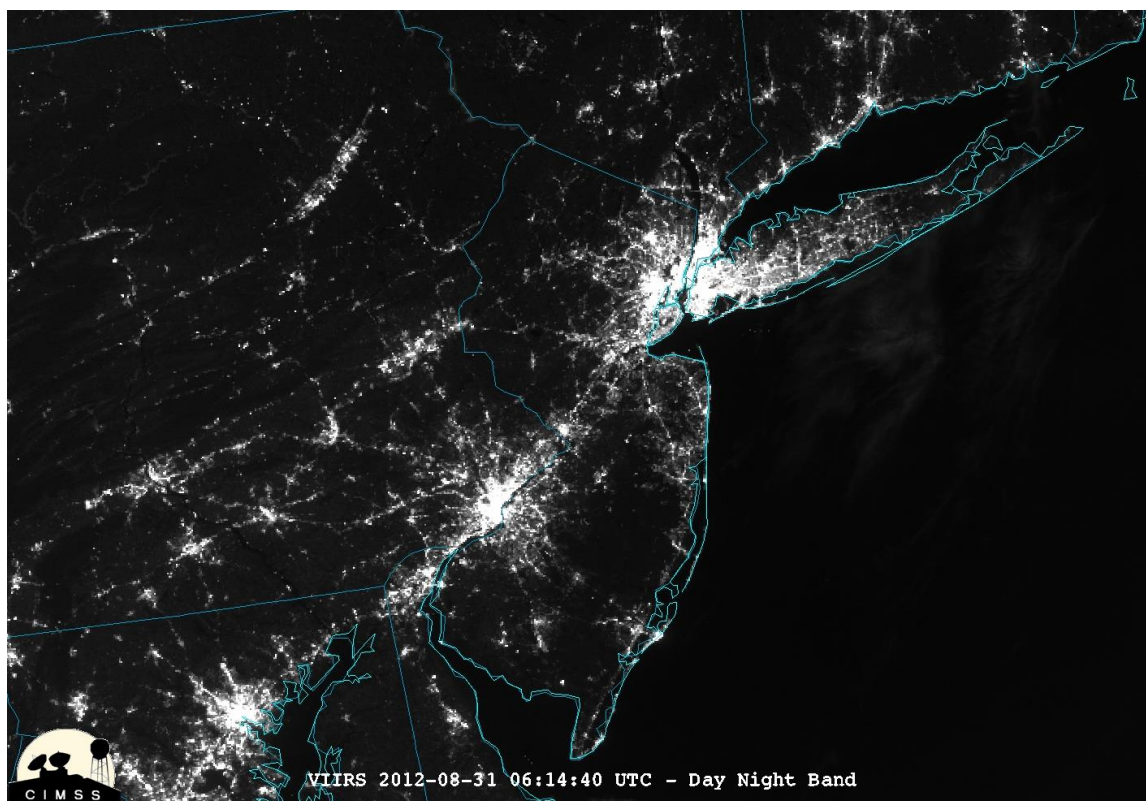


Рис. 5.12. Фрагмент космічного знімку з супутника SUOMI

Оскільки порівняння проводилось для даних з річним кроком динаміки в якості вхідних даних було використано річні усереднені композити нічних знімків, сезонні знімки для таких факторів як тепловий слід та сільськогосподарські землі, знімки з відсутністю хмарності для урбанізованих територій.

З метою оптимізації території України для навчання нейронної мережі було використано базову сітку комірок розміром 25 x 25 км., на яку було поділено всю територію України.

Кожна комірка отримала своє значення індексу економічного розвитку базуючись на статистичній залежності між показниками регіонального розвитку.

Було використано глобальну регресійну модель для перерахунку показників та встановлено нелінійну залежність:

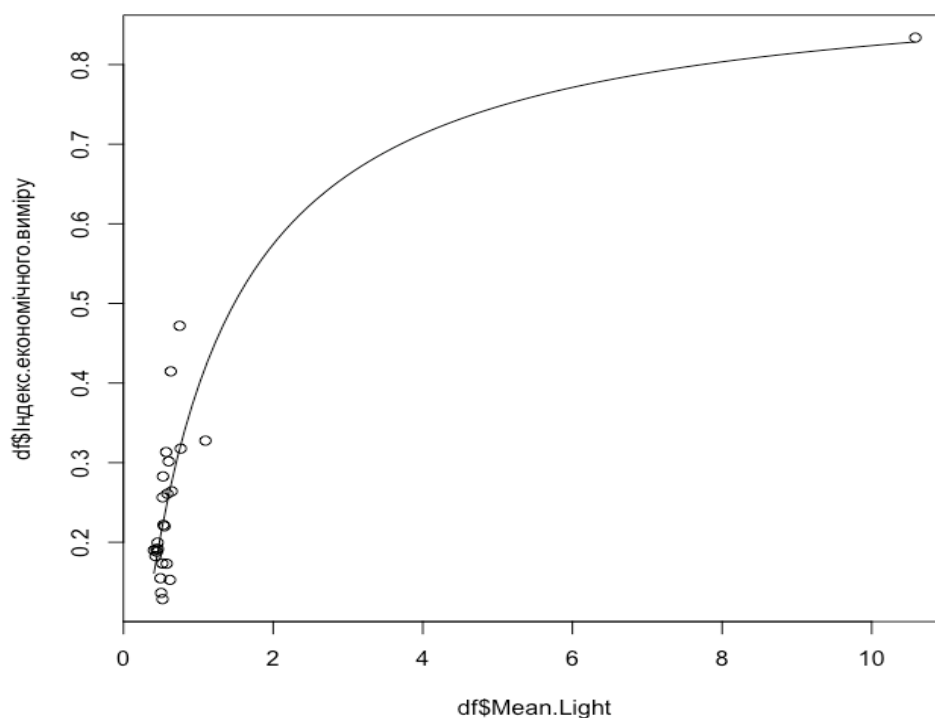


Рис.5.13. Регресійна залежність між індексом економічного розвитку та середньою

$$Y = -0.988/(x+0,9042)+0,9144, \quad (5.12)$$

Середня помилка функції 7,1 %

Таким чином, для кожної комірки було визначено значення для теплового сліду, частки сільськогосподарських земель та урбанізованих індустриальних територій. Залежності між різними типами територій та індексом економічного виміру відображено на рис. 5.14.

Це дозволило розділити вибірку на навчальну та тестову частину для проведення навчання нейронної мережі.

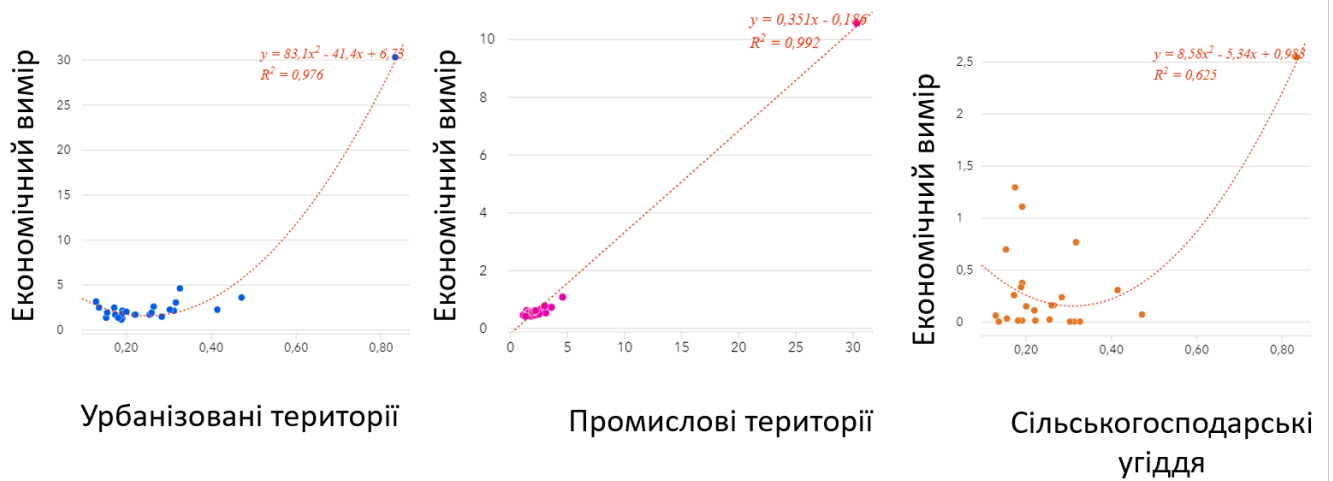


Рис. 5.14. Оцінка зв'язків між показниками використання території та економічним виміром

Біологічний нейрон імітується у ШНМ через активаційну функцію. У задачах класифікації активаційна функція повинна мати характеристику «вмикача». Іншими словами, якщо вхід більше, ніж деяке значення, то вихід повинен змінювати стан, наприклад з 0 на 1 або -1 на 1. Це імітує "включення" біологічного нейрону. У якості активаційної функції використано сигмоїдну функцію:

$$f(z) = \frac{1}{1 + \exp(-z)} \quad , \quad (5.13)$$

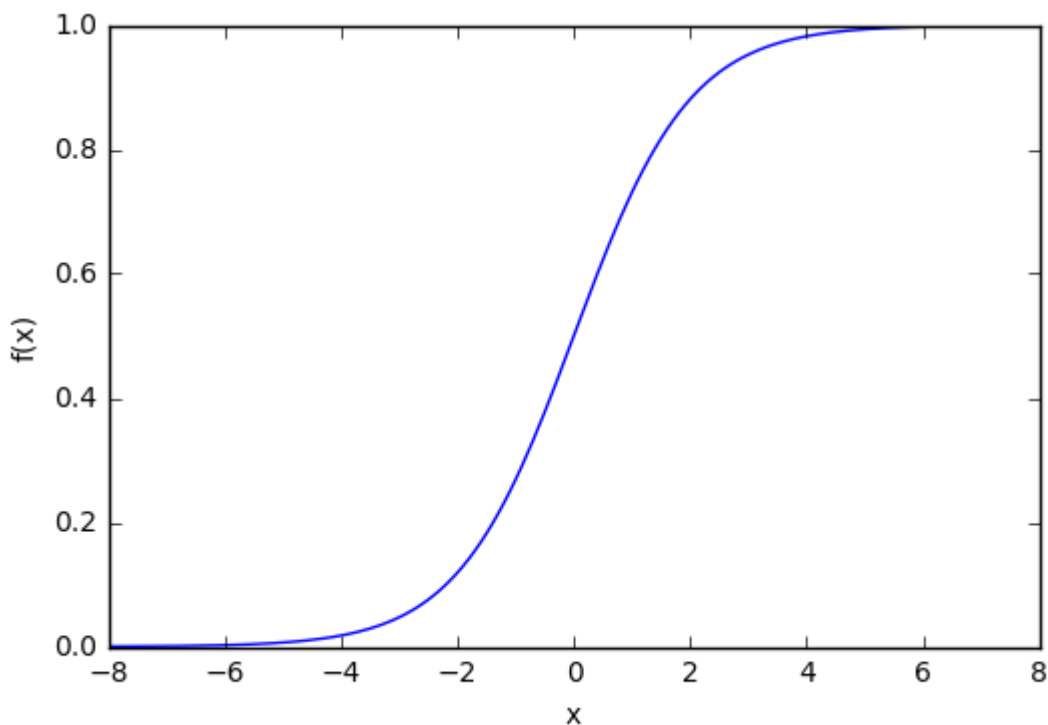


Рис. 5.15. Сигмоїдна функція активації.

$$f_{ReLU} = \max(0, x)$$

$$f_{softmax} = y_i^l = f(x_i^l) = \frac{e^{x_i^l}}{\sum_{k=0}^n e^{x_k^l}}$$

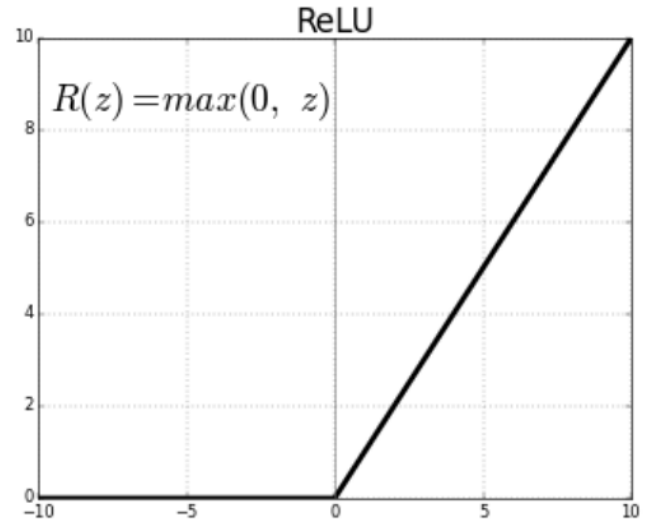


Рис. 5.16. Функція активації ReLU.

З графіку можна побачити, що функція "активаційна" – вона росте з 0 до 1 з кожним збільшенням значення x . Сигмоїдна функція є гладкою і неперервною. Це означає, що функція має похідну, що у свою чергу є дуже важливим фактором для навчання алгоритму.

Нейронна мережа для розпізнавання образів визначається набором вхідних нейронів, які можуть бути активовані за допомогою пікселів вхідного зображення. Штучні нейронні мережі - це прості математичні моделі, які визначають функцію $f: X \rightarrow Y$ або розподіл по X і Y . Математично функція нейрона мережі $f(x)$ визначається як композиція інших функцій $g_i(x)$. Широко використовується такий тип композиції, як нелінійна зважена сума:

$$f(x) = K(Z_i w_i g_i(x)),$$

де K - функція активації, Z_i - безліч значень або градація критерію $f(x)$ («шкала» критерію), набір функцій $g_i(x)$ визначається, як вектор $g = (g_1, g_2, \dots, g_n)$, а w_i - ваги, призначені для обліку важливості критеріїв.

Існує також випадкова величина $F = f(x)$, яка залежить від випадкової величини $G = g(H)$. $G = g(H)$, в свою чергу, залежить від величини $H = h(X)$ і в підсумку все величини виявляються залежні від випадкової змінної X .

У згорткової нейронної мережі, крім функції активації, також використовуються формули функціонування нейронів в згортковому і підвиборчому шарах. Формула функціонування нейрона згорткового шару мережі:

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} x^{((i-1)+s,(j+t))},$$

де $y_k^{(i,j)}$ - нейрон k -й площині згорткового шару, b_k - нейронне зміщення k -й площині згорткового шару, K - розмір рецептивної області нейрона, $w_{k,s,t}$ - матриця синаптичних коефіцієнтів, x - виходи нейронів попереднього шару.

Нейронні мережі не програмуються в звичному сенсі слова, вони навчаються. Одне з найважливіших переваг нейронних мереж перед іншими алгоритмами розпізнавання - можливість навчання. Це навчання полягає в знаходженні коефіцієнтів зв'язків між нейронами. У згортковій нейронній мережі як способі навчання, використовується алгоритм зворотного поширення помилки. Для вимірювання якості розпізнавання використовується функція середньоквадратичної помилки:

$$y_k^{(i,j)} = b_k + \frac{1}{4} w_k \sum_{s=1}^2 \sum_{t=1}^2 x^{((i-1)+s,(j+t))}.$$

де E_p - величина функції помилки способу p , t_{pj} - бажаний вихід нейрона j для образу p , o_{pj} - дійсний вихід нейрона j для образу p .

Підсумкова корекція синаптичних коефіцієнтів проходить за такою формулою:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj}.$$

де η - це коефіцієнт пропорційності, який впливає на швидкість навчання.

В ході навчання нейронна мережа може визначати складні залежності між вхідними та вихідними даними, і виконувати узагальнення. В результаті успішного навчання, нейронна мережа зможе дати вірний результат навіть при неповних або

перекручених даних. Для успішного навчання мережі потрібно наблизити вихід мережі до бажаного виходу, а щоб зменшити величину функції помилки, слід налаштовувати ваги нейронів.

Навчання нейронних мереж полягає в послідовному коригуванні синаптичних ваг між нейронами. Одним з найбільш розповсюджених і ефективних алгоритмів навчання для нейронних мереж є алгоритм зворотного поширення помилки. Свою назву алгоритм отримав завдяки тому, що помилка, яка обчислюється на кожній ітерації, поширюється по нейронній мережі від виходу до входу з метою переналаштування синаптичних ваг. В процесі навчання мережі, при подачі вхідного вектора, вихід мережі порівнюється з виходом з навчальної вибірки, формуючи помилку:

$$E_k = \frac{1}{2} \sum_{j=0}^N (t_{kj} - x_{kj})^2 ,$$

де t_{kj} - бажане значення виходу нейрона; x_{kj} - поточне значення виходу нейрона.

Коригування синаптичних ваг виконується за такою формулою:

$$\Delta w_{ij} = -\eta \delta_{kj} x_{kj} ,$$

де η - коефіцієнт швидкості навчання; x_{kj} - значення входу нейрона; δ_{kj} - помилка нейрона.

Значення помилки нейронів мережі визначається за формулою

$$\delta_i^{(q)} = (f_i^{(q)}(S))' \sum_j w_{ij} \delta_i^{(q+1)} ,$$

де $\delta_i^{(q)}$ - значення помилки i -го нейрона в шарі q ; $\delta_i^{(q+1)}$ - значення помилки j -го нейрона в шарі $q+1$; w_{ij} - вага зв'язку, що з'єднує два нейрони; $(f_i^{(q)}(S))'$ - значення похідної активаційної функції i -го нейрону в шарі q .

Перевагами даного алгоритму навчання є: простота реалізації, можливість використовувати безліч функцій втрат, можливість застосування при великих обсягах даних. До недоліків алгоритму можна віднести мале корегування ваг, що

веде до довгого процесу навчання. При цьому виникає завдання вибору оптимального розміру кроку. Занадто маленький розмір кроку призводить до повільної збіжності алгоритму, занадто великий розмір кроку може призвести до втрати стійкості процесу навчання.

Для вирішення даних проблем існують різні методи оптимізації даного алгоритму. З існуючих методів оптимізації для навчання і подальшого порівняння їх роботи було обрано такі:

- Nesterov accelerated gradient - метод оптимізації, що засновано на ідеї накопичення імпульсу. При тривалому русі в одному напрямі швидкість буде зберігатися певний час. Для цього необхідно зберігати кілька попередніх значень параметрів і обчислювати середнє значення, що займає надто багато пам'яті.

- AdaGrad (Adaptive gradient) - алгоритм оптимізації, принцип якого полягає в збереженні частоти зміни градієнта. Ідея полягає в збереженні суми квадратів оновлення для кожного параметра мережі. На основі цієї величини регулюються оновлення параметрів – параметри, що часто зустрічаються оновлюються рідше, ніж рідкі, що звільняє місце, тим самим забезпечується адаптивна швидкість навчання мережі [17].

- Adam (Adaptive moment estimation) - алгоритм оптимізації, що поєднує в собі принципи накопичення імпульсу і збереження частоти зміни градієнта. Даний метод має переваги обох розглянутих вище методів.

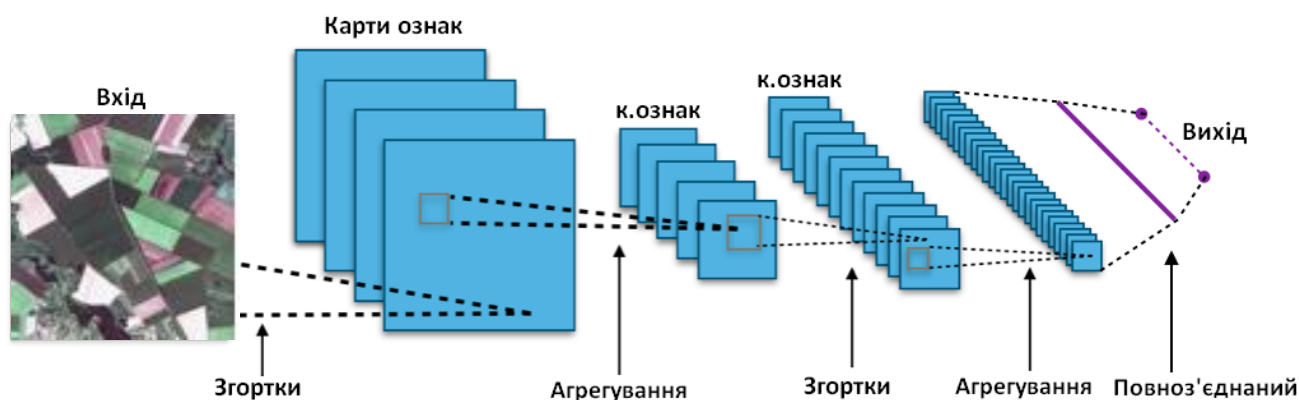


Рис. 5.17. Навчання нейронної мережі на основі даних ДЗЗ

Для навчання було використано архітектуру мережі U-Net, як одного із зразків архітектур CNN для задач сегментації зображень, коли потрібно не тільки визначити клас зображення цілком, але і сегментувати його області за класом, тобто створити маску, яка буде розділяти зображення на кілька класів. Архітектура складається з шляху звуження для захоплення контексту і шляху симетричного розширення, який дозволяє здійснити точну локалізацію.

Для U-Net характерно:

- досягнення високих результатів в задачах сегментації космічних зображень;
- використання невеликої кількості даних для досягнення надійних результатів.

Шлях звуження – типова архітектура згорткової нейронної мережі. Він складається з повторного застосування двох згорток 3×3 , за якими слідують Ініт ReLU і операція максимального об'єднання (2×2 ступеню 2) для зниження розрізнення.

На кожному етапі понижувальної дискретизації канали властивостей подвоюються. Кожен крок у шляху розширення складається з операції підвищення дискретизації карти властивостей, за якою слідують:

- згортка 2×2 , яка зменшує кількість каналів властивостей;
- об'єднання з відповідним чином обрізаної картою властивостей зі шляху звуження;
- дві 3×3 згортки, за якими слідує ReLU.

На останньому шарі використовується згортка 1×1 для зіставлення кожного 64-компонентного вектора властивостей з бажаною кількістю класів. Всього мережа містить 23 шари згортки.

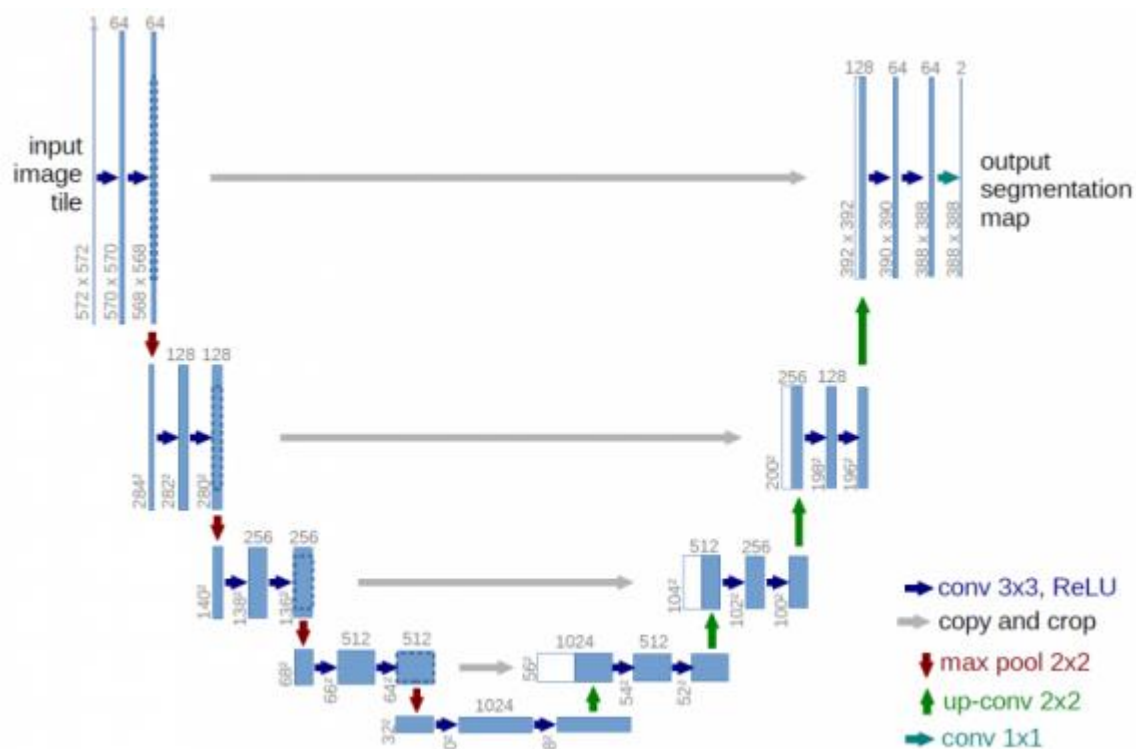


Рис. 5.18. Архітектура нейронної мережі U-net

Мережа навчається методом стохастичного градієнтного спуску на основі вхідних зображень і відповідних їм карт сегментації. Через згортку вихідне зображення менше вхідного сигналу на постійну ширину смуги. Функція softmax, яка застосовується попіксельно, обчислює енергію за остаточною картою властивостей разом з функцією крос-ентропії.

Оцінка точності навчання мережі відображена на рис. 5.19.

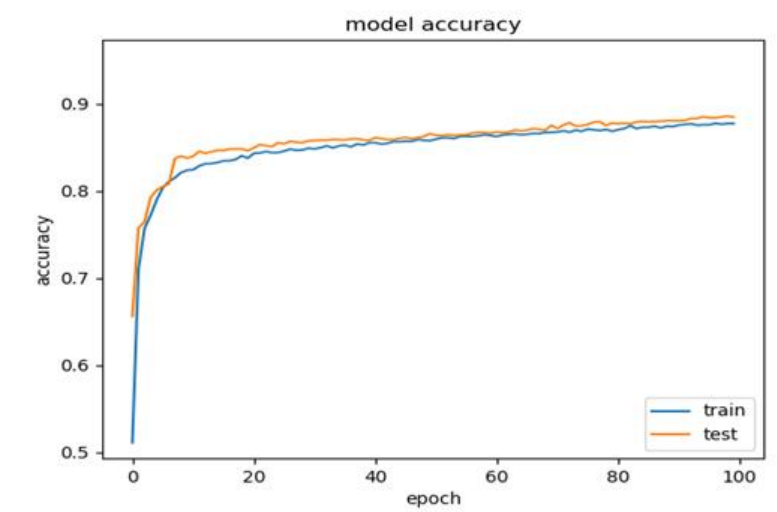


Рис. 5.19. Оцінка надійності навчання мережі

Точність навчання – 87%

Середня помилка тестової вибірки 12 %

Навчальна вибірка– 800 полігонів

Тестовая вибірка – 200 полігонів

Полігони обрані випадковим чином.

Отримана сегментація зображень за класами використання земель дозволила обрахувати значення індексу економічного розвитку території за рівняннями регресійних зв'язків відповідно до розподілу площ класів у кожній комірці сітки (рис. 5.20).

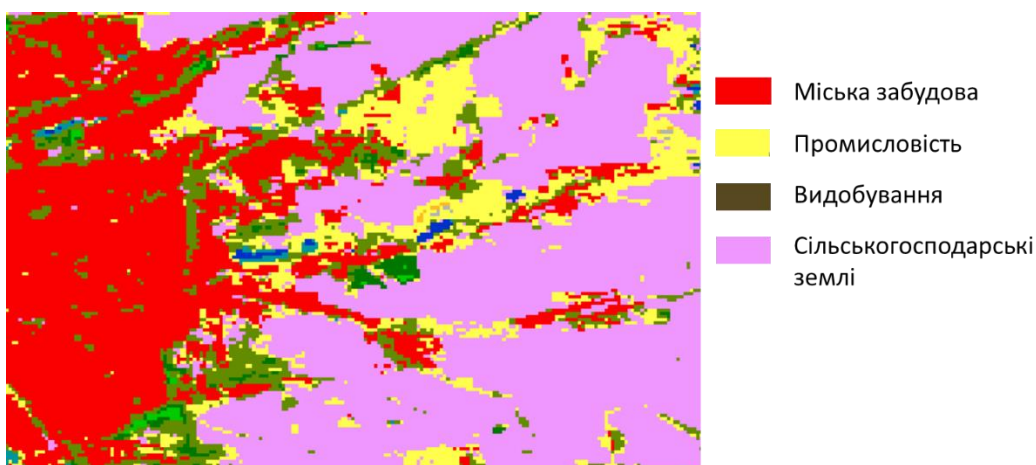


Рис. 5.20. Сегментація зображень за класами використання земель.

Це дозволило отримати дані індексу економічного розвитку для всієї території України за регулярною мережею комірок.

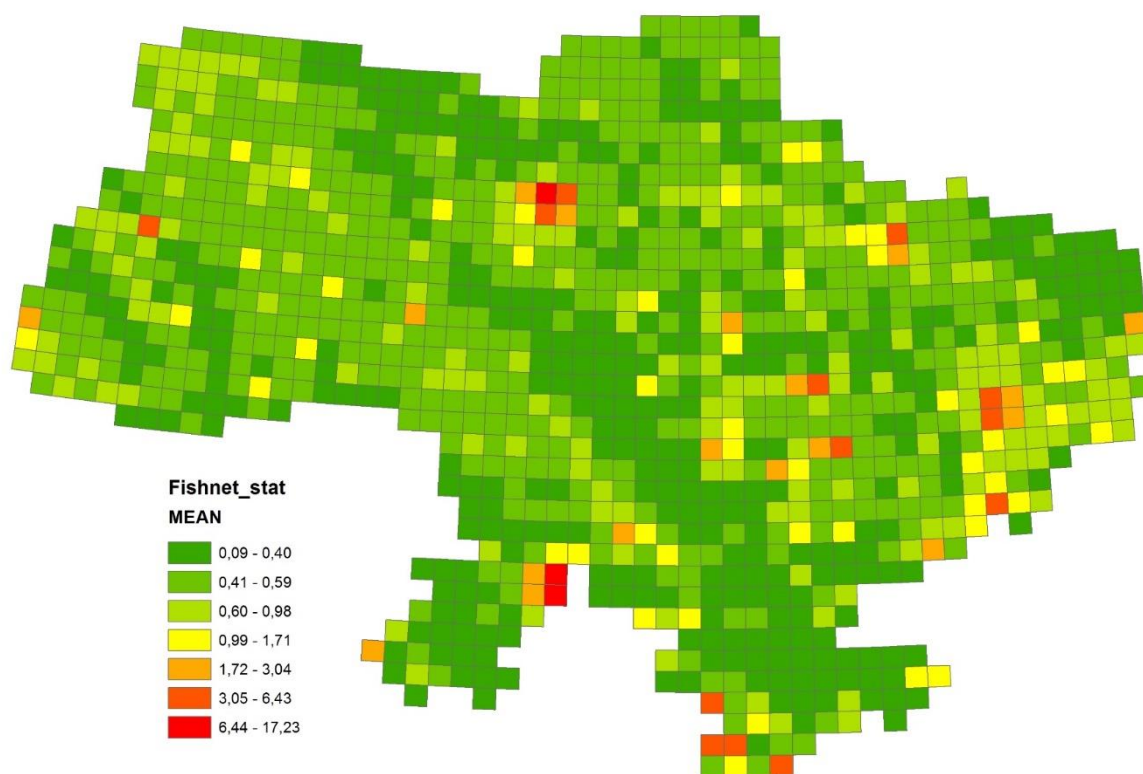


Рис. 5.21. Розраховані показники індексу економічного виміру в 2014 р.

З метою переходу від регулярної мережі комірок (D) до одиниць адміністративно-територіального устрою (A) території України (R) було запропоновано розрахунок зональної статистики на базі правил кон'юнкції:

$$D_n \cup A_n \in R.$$

В результаті було отримано значення індексу економічного розвитку для адміністративних одиниць рівня район та місто обласного значення.

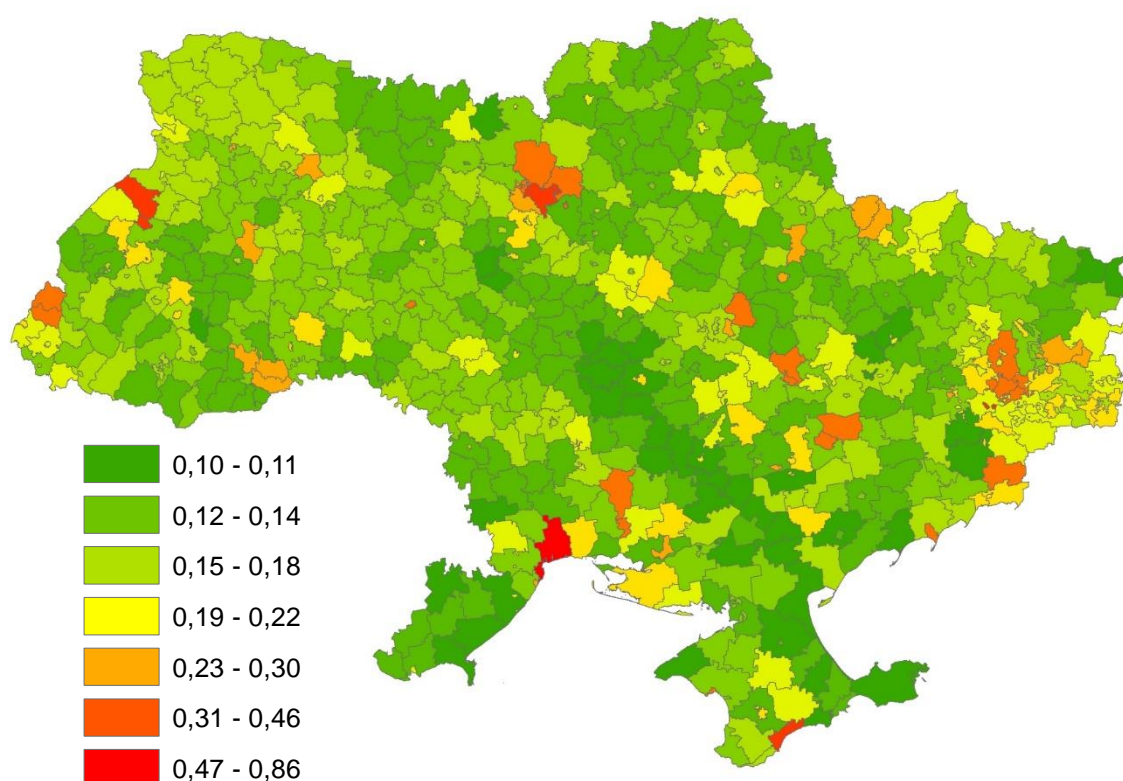


Рис. 5.22. Значення індексу I_e за районами та містами обласного значення

Значення закомірками було обраховано для трьох років 2014, 2015, 2016 рр. Це дає змогу отримати інформацію про стан економічного розвитку окупованих територій та АР Крим, яка на сьогодні є недоступною.

$$I_{\Delta} = I_{2016} - I_{2014} . \quad (5.29)$$

Таким чином, було отримано характеристику зміни динаміки індексу економічного розвитку за період 2014-2016 рр. I_{Δ} змінюється в діапазоні від -1 до 1.

Найбільше зростання за цей час показують великі міста Київ, Харків, Дніпропетровськ. Певні позитивні тенденції притаманні південним територіям України, де розвивається сільське господарство. Регіон Донбасу та АР Крим характеризується зниженням економічного розвитку у порівнянні з 2014 р.

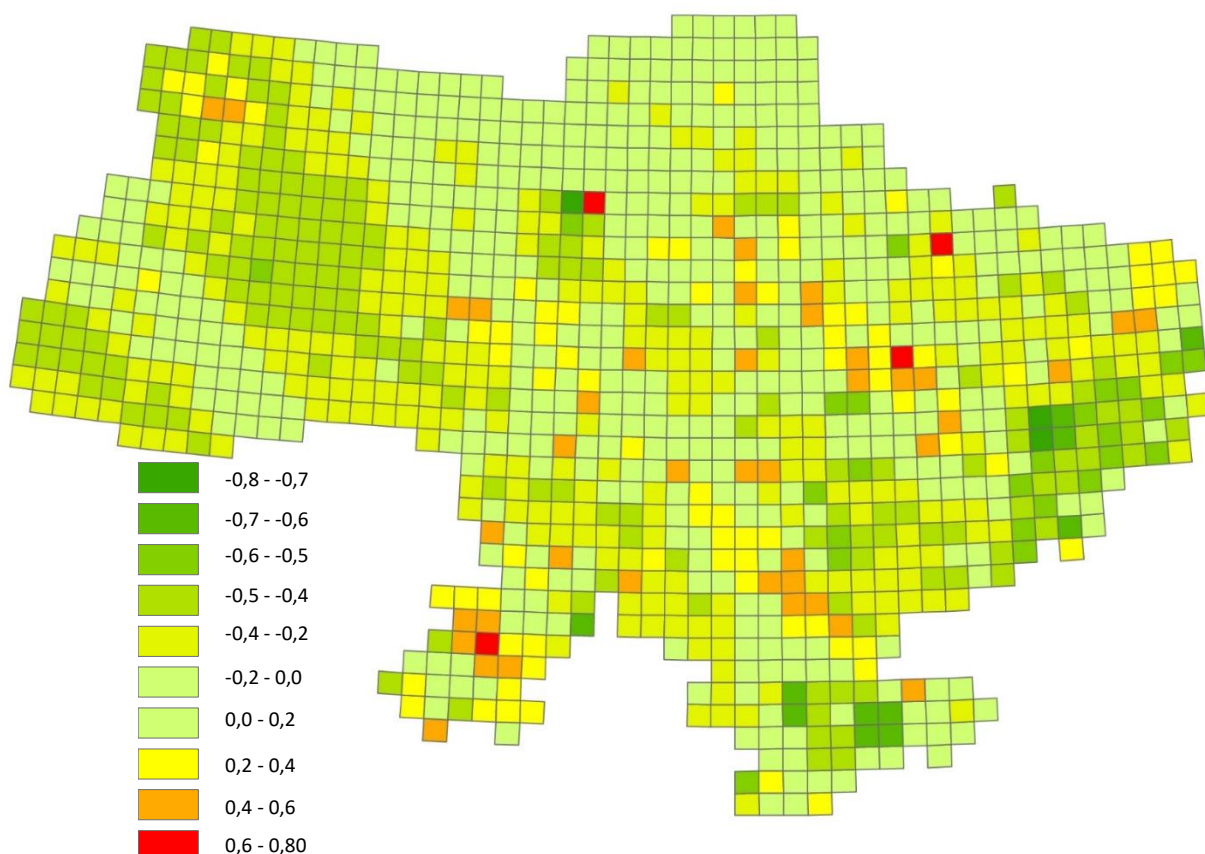


Рис. 5.23. Зміна індексу економічного виміру 2014 – 2016

В результаті можна зазначити, що використання методу багатошарових нейронних мереж можуть дозволити отримати кількісні дані, що характеризують процес сталого розвитку території незалежно від статистичних джерел інформації з високим ступенем надійності. Основними властивостями цих даних є розрізнення вхідних космічних даних, крок мережі для апроксимації інформації та метод навчання нейронної мережі.

5.5. Реалізація елементів СППР інтелектуального аналізу геопросторових даних для задач сталого розвитку

Система підтримки прийняття рішень на основі методів інтерелектуального аналізу геопросторових даних складається з наступних основних блоків (рис. 5.24).

- Блок постановки задач
- Блок управління даними
- Блок аналізу та моделювання
- Блок оцінювання результатів та прийняття рішень

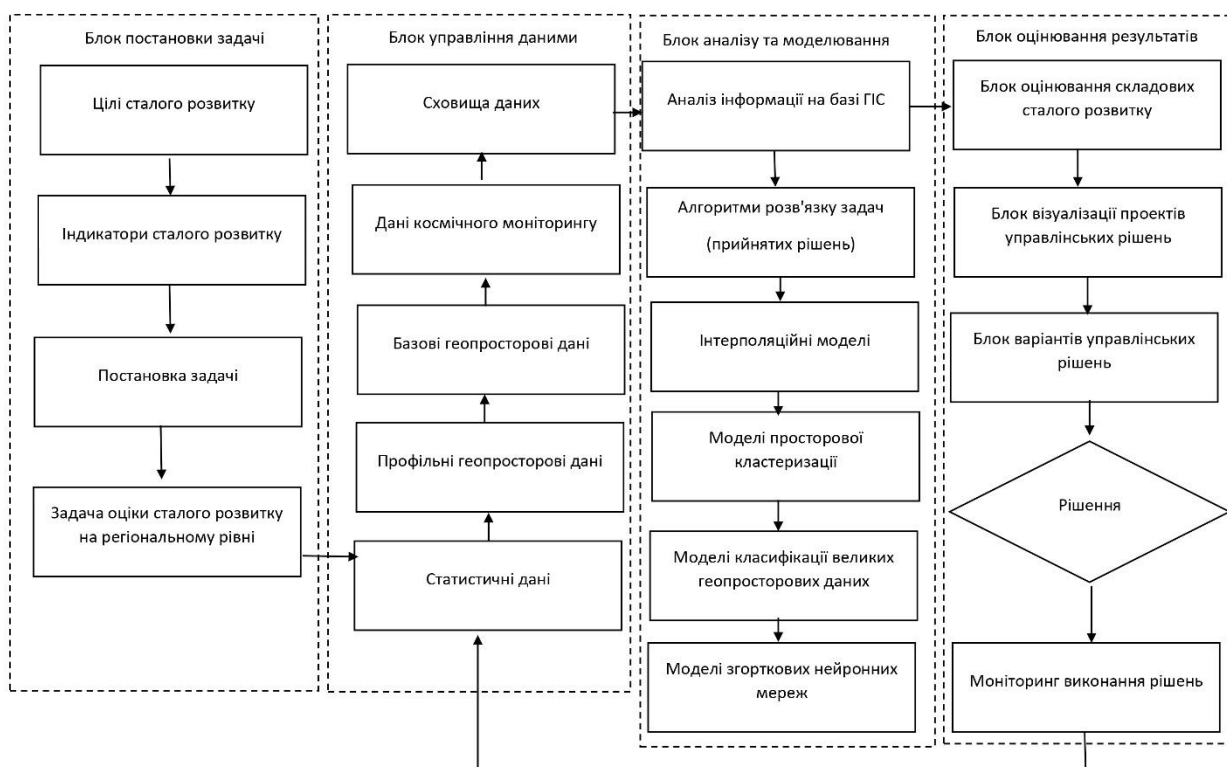


Рис. 5.24. Загальна архітектура СППР на основі методології інтелектуального аналізу в задачах сталого розвитку

Блок постановки задач ґрунтується на методології сталого розвитку, яка була розвинута за останні 30 років. В основу цієї методології покладено управління досягненням визначених цілей сталого розвитку. Загалом, виділяється 17 основних цілей, які мають свою ієрархію реалізації. На основі сукупності цілей побудовано математичну модель індикаторів сталого розвитку, яка оцінює якість життя та безпеку населення. Постановка задачі формується на основі прогнозування складових сталого розвитку на основі геопросторових даних обраними методичними підходами до інтелектуального аналізу даних на регіональному рівні.

Блок управління даними формується на основі використання даних з просторових та непросторових баз даних. Основні типи геопросторових даних представлені векторними та растровими моделями. До основних видів векторних моделей відноситься базові та профільні набори даних, що містять інформацію у форматі точкових об'єктів, ліній та полігонів про адміністративно-територіальний устрій та локалізацію подій. Растрові набори даних можуть містити результати

інтерполяції точкових поверхонь у форматі GRID, космічні зображення та результати їх класифікації. Дані у растровому форматі можуть зберігатися у базі геоданих або у промисловій реляційній СУБД. Структура збереження результатів космічного знімання може бути організована у вигляді окремих каналів, синтезованих зображень, мозаїки.

Обробка та аналіз даних відбувається в блоці аналізу та моделювання на базі компонентів розподіленої веб-орієнтованої геоінформаційної системи, що призначена для управління та обробки даних з геопросторовою прив'язкою. Для цілей аналізу використовуються як окремі методи ГІС-аналізу, методи інтелектуального аналізу даних, а також ансамблі методів, які поєднують ці групи методів з методами системного аналізу. До основних груп методів відносяться методи геостатистичного аналізу на базі інтерполяційних методів, які дозволяють обчислити відсутні значення, підвищити дискретність моделі та прогнозувати зміни, методи геопросторової класифікації, кластеризації та районування, які дозволяють виділити однорідні або схожі групи адміністративних одиниць для прийняття управлінських рішень, моделі обробки великих обсягів даних для безпекової компоненти території, моделі згорткових нейронних мереж для класифікації космічних супутникових знімків з метою оцінювання індикаторів сталого розвитку.

Блок оцінювання результатів та прийняття управлінських рішень містить модуль оцінювання складових сталого розвитку на основі результатів обробки геопросторових даних та математичної моделі. В залежності від абсолютних показників індексів та співвідношення між ними приймаються варіанти управлінських рішень, які пропонуються до впровадження. Для підтримки прийняття рішень використовується блок візуалізації даних у вигляді електронних картографічних представлень та інтерактивних дашбордів. СППР передбачає проведення моніторингу прийнятих рішень та їх подальшої адаптації в разі необхідності.

СППР розроблена на базі серверної платформи ArcGIS Enterprise 10.6.

ArcGIS Enterprise є базовою системою програмного забезпечення для ГІС, що забезпечує картографування і візуалізацію, аналітику і управління даними. Це основа для запуску набору додатків компанії Esri і власних користувальницьких додатків. ArcGIS Enterprise тісно інтегрований з ArcGIS Desktop і ArcGIS Pro для картографування і створення карт, а також легко підключається до ArcGIS Online для публікації ресурсів між системами.

ArcGIS Enterprise надає повний контроль над розгортанням. Його можна запустити на одному сервері або масштабувати на декількох машинах: на локальному ресурсі, на фізичному обладнанні або у віртуальному середовищі

Ця гнучкість відкриває розширені можливості для підтримки прийняття рішень, таких як відстеження даних в реальному часі, виконання аналізу великих даних, растрова аналітика, розміщення сервісів зображень і виконання робочих процесів вивчення даних (рис. 5.25).

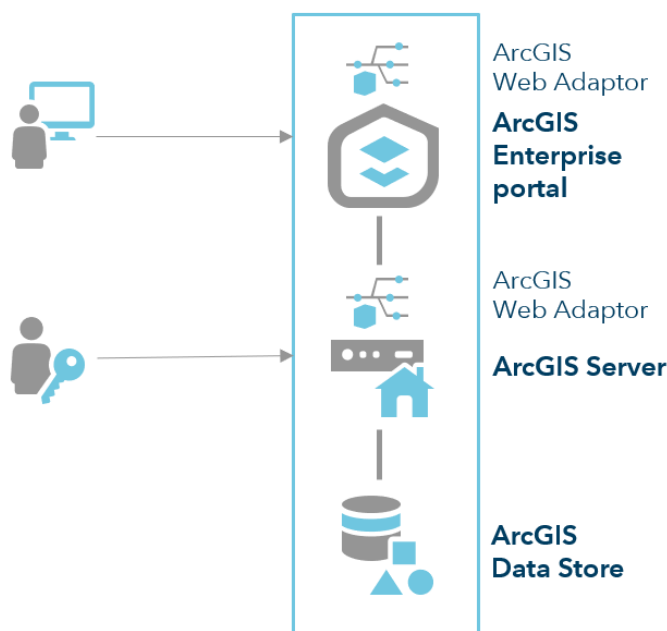


Рис. 5.25. Архітектура складових компонентів ArcGIS Enterprise 10.6.

Чотири компонента ArcGIS Enterprise працюють разом, забезпечуючи комплексну функціональність для веб-картографування, роботи з зображеннями, обробки даних в режимі реального часу, пакетного аналізу великих обсягів даних і наукових просторових даних:

- ArcGIS Server підтримує створення карт і аналіз в ГІС і динамічно масштабується для підтримки трафіку.
- Портал ArcGIS Enterprise - це центральний вузол, де користувачі створюють, публікують і управляють картами, додатками і просторовими даними, а також спільно використовують їх з колегами.
- ArcGIS Data Store забезпечує гнучке сховище даних для хост-сервера, що використовується в розгортанні.
- ArcGIS Web Adaptor об'єднує ArcGIS Server і Portal for ArcGIS з вже існуючими веб-серверами і інфраструктурою безпеки організації.

Додатковий модуль ArcGIS Image Server служить для публікації сервісів зображень з метою візуалізації в реальному часі і виконання розподіленого аналізу на растрових даних. ArcGIS GeoAnalytics Server виконує розподілений аналіз на великих обсягах табличних і векторних даних. ArcGIS GeoEvent Server служить для інтеграції поточкових даних подій в реальному часі з таких ресурсів, як Інтернет речей, в ГІС-систему.

Постачальники геоданих взаємодіють з єдиним сховищем просторових і атрибутивних даних. За публікацію просторових даних відповідає веб-ГІС-сервер, який є сполучною ланкою між користувачами і базою геоданих.

Користувачам, які мають стандартний браузер, не потрібно встановлювати геоінформаційне програмне забезпечення на свій комп'ютер, тому що вся робота з даними здійснюється у вікні браузера. Карта може передаватися як у векторному, так і в растровому вигляді.

Розвиток геоінформаційних систем в Інтернеті супроводжується створенням інтерактивних середовищ взаємодії клієнтських додатків з геоінформаційним сервером, зміст яких полягає в наданні можливості користувачеві самостійно вибирати на карті ділянки для відображення на своєму комп'ютері. На сервері розміщується набір серверних програм, що забезпечують взаємодію з клієнтом, аналіз дій клієнта і створення растрової картини на область, яка визначена користувачем.

Взаємодію клієнта з сервером можна представити наступною моделлю: користувач бачить генералізовану ділянку карти і, вибираючи більш дрібні ділянки, отримує все більш детальне відображення місцевості. Таке інтерактивне взаємодія вимагає більш потужних серверів.

Використання векторного підходу дозволяє отримати з програми перегляду ідентифікатор обраного об'єкта і потім запросити атрибутивну інформацію з сервера. У разі визначення об'єкта на растровому зображенні карти існує тільки один підхід - передача ГІС-серверу координат об'єкту, за якими запитується інформація (у вигляді HTTP-запиту). У цьому випадку сервер своїми засобами виділяє потрібний об'єкт, враховуючи доступність різних шарів карти та інші обмеження, і повертає відповідь клієнту по протоколу HTTP. Формат взаємодії клієнта і сервера в мережевій ГІС визначається стандартами Open GIS Consortium, Inc (<http://www.opengis.org/>), яких дотримується більшість розробників, які представляють свої продукти на ринку мережових ГІС-технологій. Відповідь повертається в форматі GML (Geography Markup Language) і підлягає додатковій обробці або вимагає визначення схеми інтерпретації відповіді браузером.

Прикладом взаємодії клієнта з таким ГІС-сервером може бути наступна технологічна послідовність. Користувачеві пропонується узагальнена (генералізована) карта, на якій він може вибрати ділянку для подальшої деталізації. Також на сторінці сервера пропонується набір функцій для пошуку інформації за атрибутивною ознакою.

Після набору необхідних даних для атрибутивного пошуку користувачеві надається або об'єкт пошуку в графічній формі, або список знайдених об'єктів. Вибираючи об'єкти, користувач отримує їх картографічне зображення і має можливість перегляду атрибутивної інформації, приєднаної до графічного об'єкту. Слід звернути увагу, що при використанні інтернет-технологій з'являється можливість інтеграції кількох різних проектів з ГІС-модулем, що значно розширює базові можливості геоінформаційної системи.

В системі з триланкової архітектурою веб-ГІС виділяють кілька основних частин.

Клієнтський додаток призначений для перегляду карти і представлений веб-браузером («тонкий» клієнт), модулем відображення даних, що вбудований в веб-браузер, або призначеним для користувача додатком («товстий» клієнт). Засоби для редагування, адміністрування та компонування карти теж працюють з сервером віддалено та відносяться до клієнтських додатків. Всі клієнтські додатки отримують від сервера дані за певним шаблоном, що вказується в запиті. Файли шаблону зберігаються на сервері і включають інформацію про налаштування шарів карти, що необхідні для її коректного відображення. Ця інформація включає імена шарів, кольори і типи ліній, шлях до джерела даних і іншу інформацію про карту.

Веб-сервер виконує функції посередника між клієнтськими і серверними додатками, який використовується для передачі запитів і даних по протоколу HTTP. Сучасний веб-сервер, що використовується для задач мережевої ГІС, дозволяє виконувати спеціальні додатки, які слугують для формування та обробки запитів, роботи з додатковими атрибутивними даними, рішення прикладних задач, розширення функціональності ГІС.

У разі мережевої ГІС на рівні веб-сервера розташовується також диспетчер запитів. Цей функціональний модуль служить для підтримки зв'язку клієнта з ГІС-серверами. Він приймає запити на картографічні дані, що приходять на веб-сервер від різних клієнтських додатків, потім їх відтворює, розподіляє і передає для обробки на ГІС-сервер. Багаторівнева архітектура і наявність такої ланки, як «диспетчер», дозволяють створити систему з декількома розподіленими ГІС-серверами.

Сервер додатків в концепції мережевої ГІС являє собою веб-ГІС-сервер. Даний компонент складається з декількох взаємодіючих частин. Основним блоком є «Виконавча служба». Цей блок отримує і обробляє запити від диспетчера, оперуючи просторовими і атрибутивними даними, що отримані з кількох джерел даних. Потім дані форматуються і надсилаються через веб-сервер на запит клієнту по протоколу HTTP. При цьому є відмінності при передачі растрових і векторних даних від сервера до клієнта.

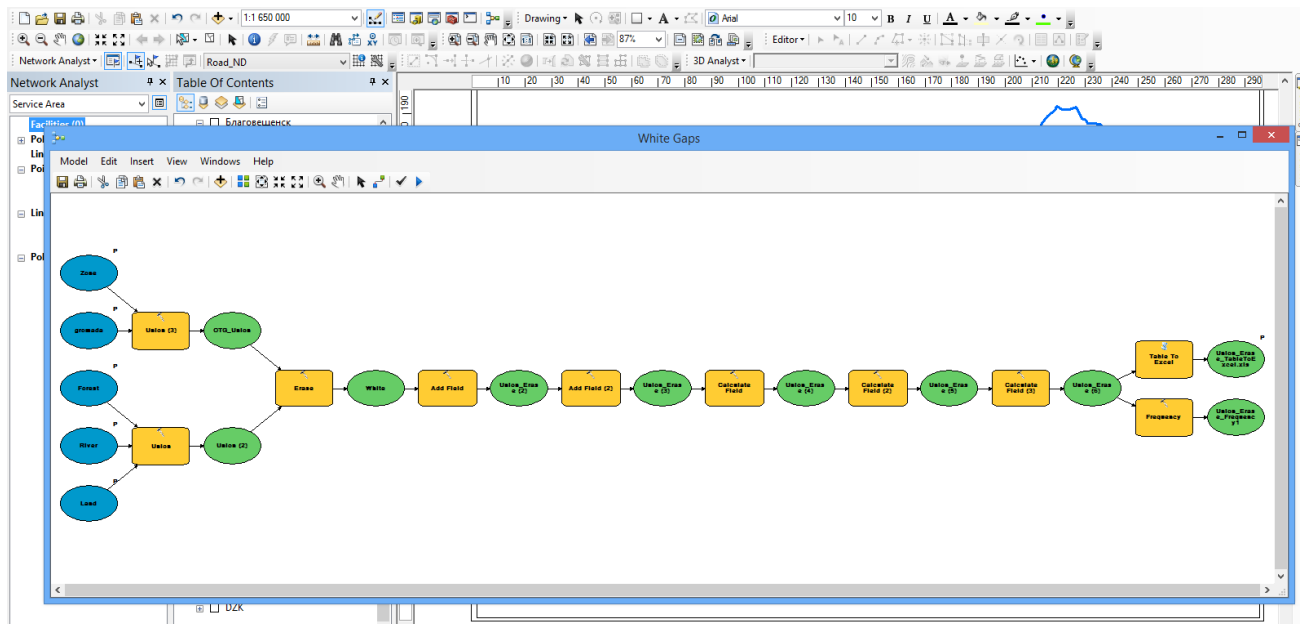
При передачі растрових даних атрибутивна інформація про об'єкти карти залишається недоступною для користувача і її необхідно або завантажувати, використовуючи додаткові ресурси, або запитувати окремо в процесі роботи з картою. Для цього до складу веб-GIS-сервера включається додатковий модуль з обробки картографічних даних для растрового варіанту карти «Служба растрового перетворення». Цей модуль реалізується у вигляді додатку сервера додатків Apache TomCat (вільно розповсюджується) і служить для формування растрового варіанту карти і повернення атрибутивної інформації. Під час отримання запиту від диспетчера цей додаток завантажує з веб-сервера файл збірки (шаблон карти), а потім робить запит даних за тим же механізмом, що і клієнтські програми (тобто за допомогою HTTP-запиту).

Після отримання даних модулем вони обробляються і вилучена атрибутивна інформація в форматі GML передається в додаток. Аналогічна схема використовується і при роботі з мобільними клієнтськими додатками, де потрібна адаптація даних під конкретний пристрій. Для отримання даних використовується механізм ODBC.

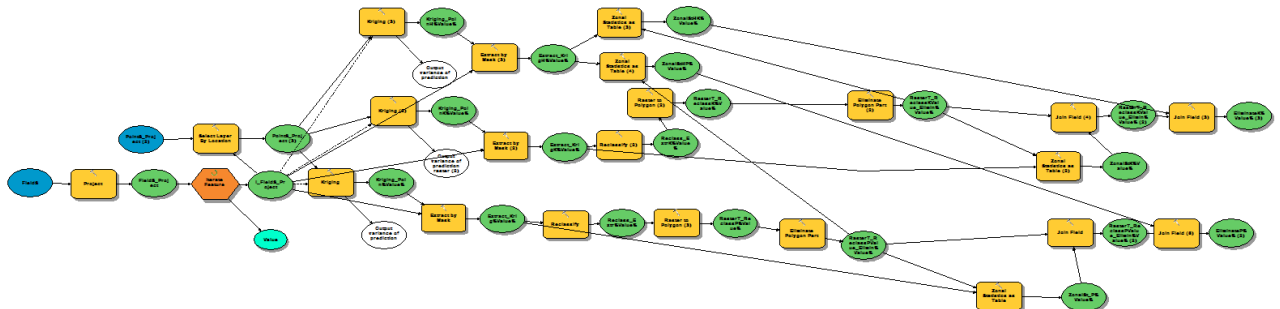
Сервер баз даних. Цей компонент вносить ще більшу варіативність в технологію, так як «Виконавча служба» може запитувати дані з віддалених серверів БД, використовуючи механізми постачальників даних. Модуль обробки просторових даних є частиною ГІС-сервера і відповідальний за завантаження всієї просторової інформації. Але найбільш правильним є просторовий компонент як частина системи управління БД (СУБД).

В системі реалізовано чотири базові модулі аналізу даних в СППР, до яких відносяться: модуль геостатистичного аналізу, модуль кластеризації та районування, модуль обробки великих даних, модуль нейромережевого аналізу з використанням архітектури згорткових мереж.

Всі модулі реалізовано на основі засобів ModelBuilder для розробки та управління інтегрованими моделями обробки геоданих з використанням мови програмування Python та подальшого їх використання у складі системи ArcGIS Enterprise 10.6 (рис. 5.26).



а)



б)

Рис. 5.26. Розробка прикладних моделей в середовищі ArcGIS ModelBuilder: а) модель кластеризації геопросторових даних; б) модель геостатистичного аналізу засобами крігінгу.

Розроблено користувацький інтерфейс модулів у вигляді майстрів покрокової обробки даних та налаштування форм з вибором джерел даних та ключових параметрів налаштування (рис. 5.27 – 5.29).

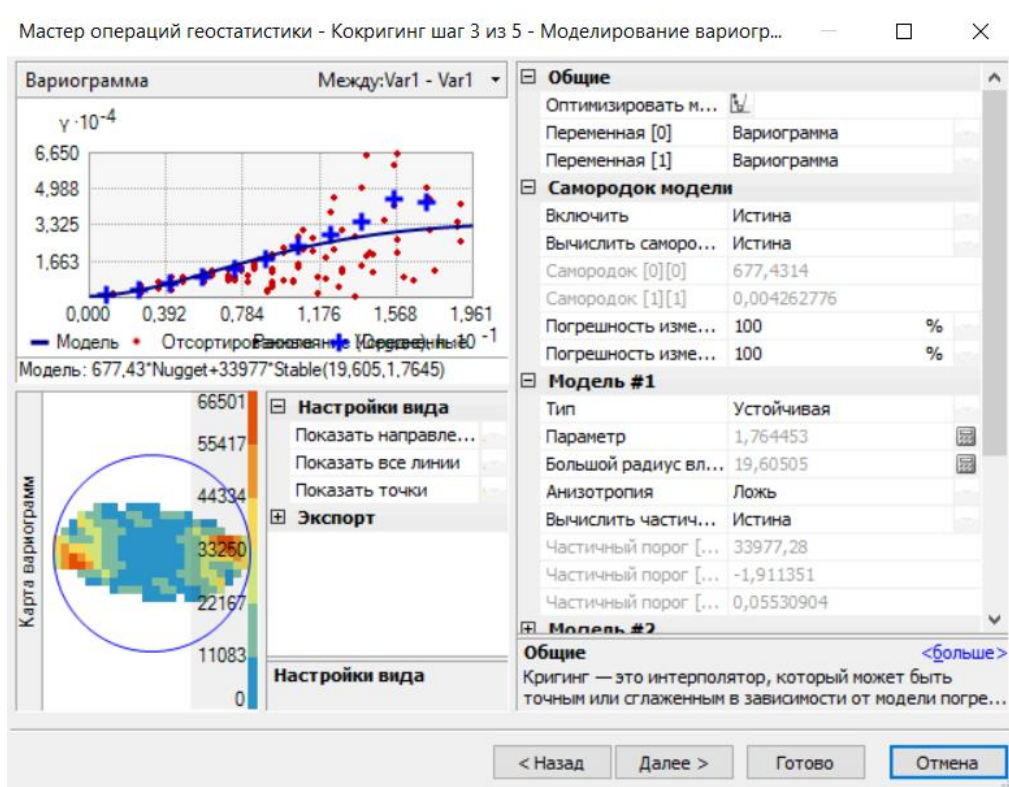


Рис. 5.27. Майстер аналізу геостатистичних моделей методом ординарного крігінгу

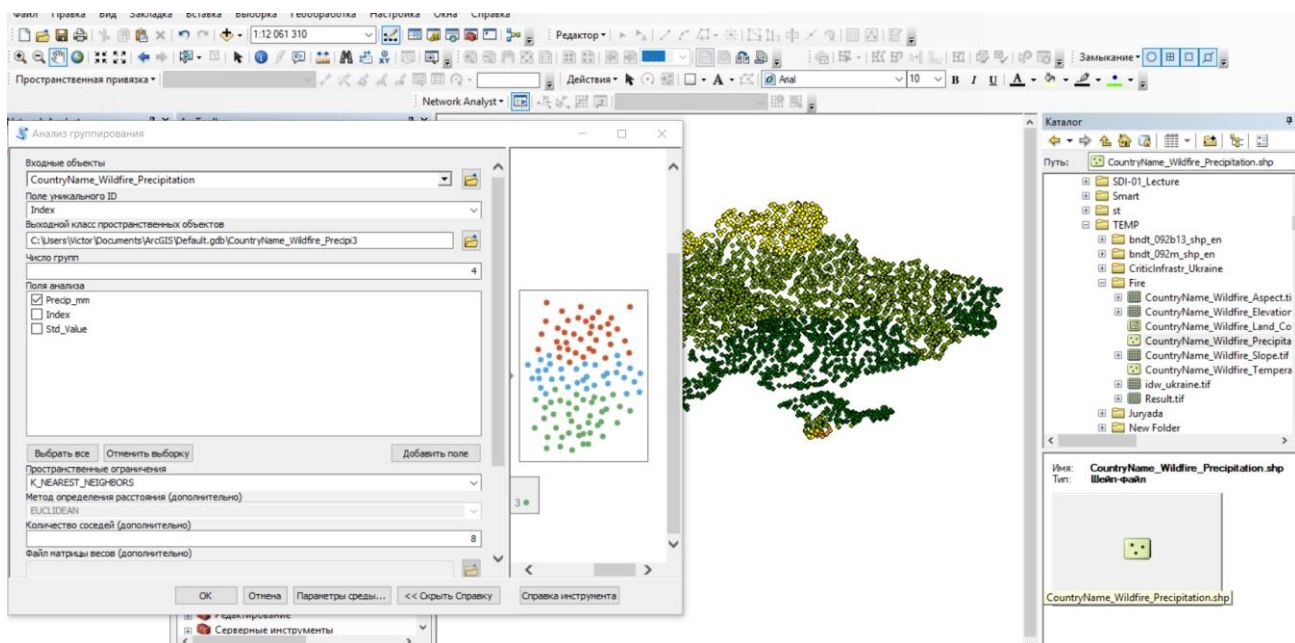


Рис. 5.28. Інтерфейс користувача у програмному комплексі ArcGIS.

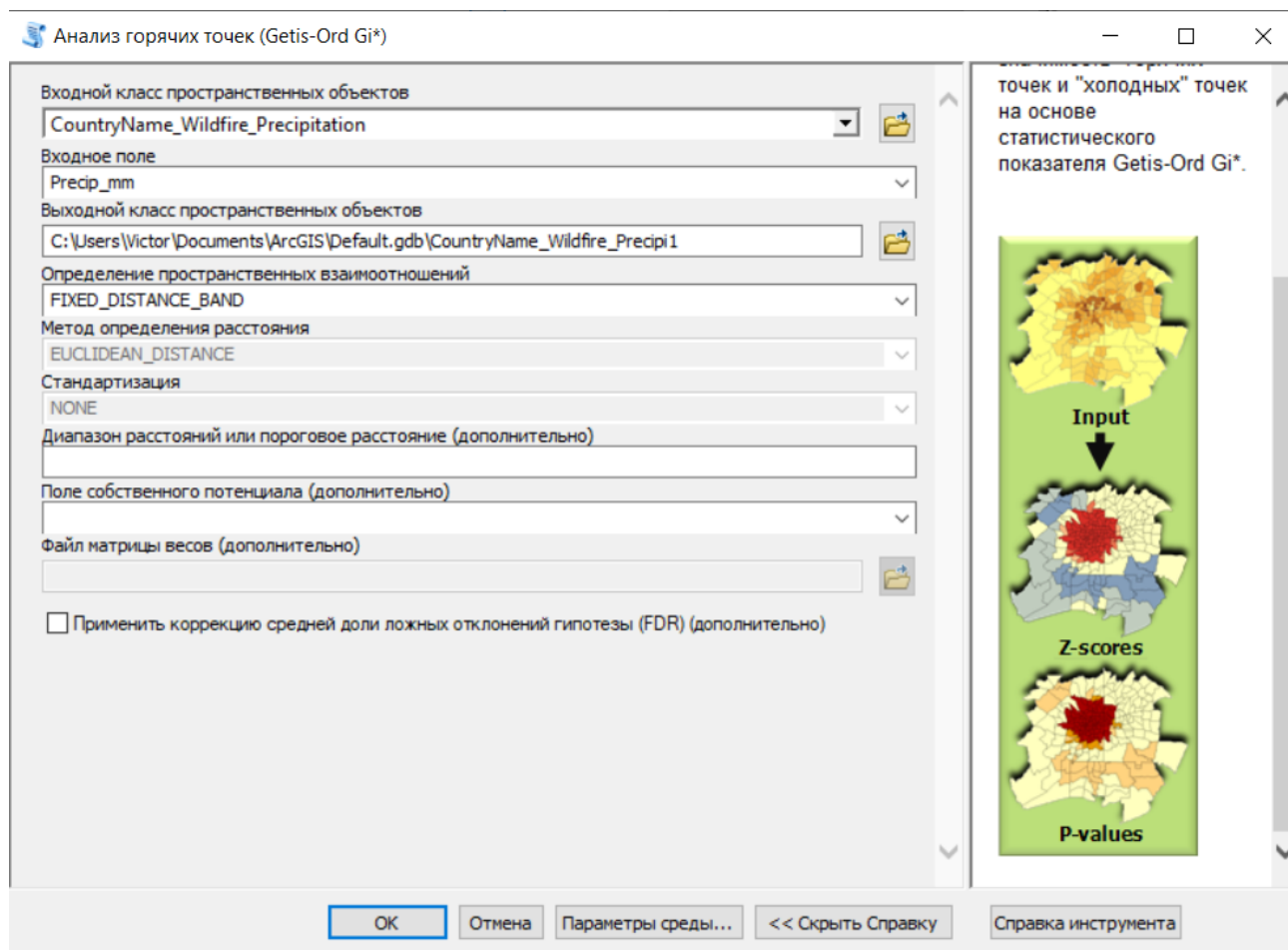


Рис. 5.29. Інтерфейсу користувача в модулі кластеризації методом аналізу гарячих точок.

Архітектурні рішення зі створення інформаційних систем для підтримки прийняття рішень повинні спиратися на просторово-часові дані проаналізованих об'єктів і базуватися на сучасних клієнт-серверних, геоінформаційних та веб-технологіях, що дозволить на якісно новому рівні задовольнити потреби оперативного управління, а також підвищить ефективність прийняття управлінських рішень.

Висновки до розділу

В математиці засоби багатовимірною просторового аналізу постійно розвиваються і добре забезпечені методами дослідження геометрії, топології та інших властивостей абстрактних об'єктів, їх множин і структур. Тому їх більш широке використання в геоінформаційних технологіях все більш актуально.

Досліджено, що нейромережеві методи в ІАГД застосовуються не тільки для аналізу даних, а також для побудови моделей процесів, що розгортаються в багатовимірних просторах.

Нейромережі, які навчаються за зразком, походять від персептронів і в сучасному трактуванні можуть розглядатися як варіанти і модифікації мереж зі зворотним поширенням помилки (іноді як результат примітивізації такого роду мереж з метою спрощення реалізації).

Визначено, що до цього класу можна віднести одношаровий і багатшаровий перцептрон, машину Больцмана, мережі, які навчаються за правилом Хебба, рекурентні шаруваті і повнозв'язні мережі зворотного поширення помилки, мережі, що використовують радіальні базисні функції.

Окремо виділено нейромережі, які навчаються без зразку - це карти Кохонена, системи з множинними локально стійкими станами, такі, як мережа Хопфілда, мережі, що налаштовуються на основі адаптивного резонансу.

Для апробації застосування нейронних мереж було використано нейронні мережі без навчання для вирішення завдання кластеризації адміністративних одиниць України за структурою землекористування та з навчанням для вирішення завдання оцінки економічного виміру індексу сталого розвитку на мікрорівні для окремих населених пунктів та територій.

В ході дослідження було розроблено схему та методику обробки даних для задач дослідження території України, яка складається з етапів предпроцесінга даних та отримання зональної статистики. На основі отриманих статистичних даних було побудовано кластери розподілу існуючих адміністративних районів України за співвідношенням основних типів земних покривів на основі використання одного з видів нейронних мереж – карт самоорганізації Кохонена.

В результаті аналізу отриманих кластерів встановлено можливість прикладного використання кластеризації за адміністративними одиницями для розробки напрямів, сценаріїв та планів розвитку території, з'ясування її сучасного стану використання та корегування цих показників у майбутньому.

Освітлення території, яке вимірюється на знімках нічного часу, є індикатором багатьох соціальних процесів, в тому числі економічного розвитку.

Світовий центр даних з геоінформатики та сталого розвитку розробив власну методику вимірювання компонентів сталого розвитку, яка включає вимірювання соціального, економічного та екологічного вимірів якості життя.

За результатами сумісного аналізу освітленості та індикаторів сталого розвитку зроблено наступні висновки.

Статистика освітленості може бути використана в якості індикатора соціально-економічного розвитку. Розрахунок індексу економічного виміру має достатньо високе значення кореляції з освітленістю (0,83), що підтверджує тісний зв'язок між цими показниками. Найбільше значення кореляції мають показники регіональної освітленості з загрозою «Вплив на змінювання клімату». Таким чином, статистика освітленості може використовуватися також для оцінки емісії вуглецю.

Наявність тісного зв'язку між ступенем освітлення території у нічний час та індексом економічного розвитку дозволило побудувати функціональну залежність між показниками та отримати навчальний набір даних для нейронної мережі.

Матриця кореляційно-регресійних зв'язків дозволила виявити зв'язки яскравості освітлення та економічного виміру сталого розвитку з урбанізованими територіями, в тому числі промисловими та житловими зонами, а також сільськогосподарськими територіями.

Було застосовано навчання згорткової нейронної мережі з градієнтним спуском на основі мультиспектральних зображень та знімків нічного освітлення, які дозволили з високою надійністю розпізнавати на знімках території міських та сільських населених пунктів, промислових зон та сільськогосподарських угідь. Під час навчання було застосовано різні варіанти функцій активації згорткової мережі. Найкращий результат було досягнуто з використанням функції softmax. Надійність навчання мережі склала 87%. Це дозволило вперше оцінити значення індексу виміру економічного розвитку для окупованих територій України та АР Крим, сформувати розподіл значень індексу для адміністративних одиниць рівня району

та міста, а також визначити динаміку індексу за останні роки з використанням інформації з космічних знімків.

В результаті можна зазначити, що використання методу згорткових нейронних мереж може дозволити отримати кількісні дані, що характеризують процес сталого розвитку території незалежно від статистичних джерел інформації з високим ступенем надійності. Основними властивостями цих даних є розрізнення вхідних космічних даних, крок мережі для апроксимації інформації та метод навчання нейронної мережі.

Розроблено загальну архітектуру СППР на основі методології інтелектуального аналізу в задачах сталого розвитку, яка включає блоки постановки задач, управління даними, аналізу та моделювання, оцінювання результатів та прийняття рішень.

СППР реалізовано на основі веб-орієнтованої розподіленої геоінформаційної системи ArcGIS Enterprise 10.6. шляхом розробки аналітичних модулів геостатистичного аналізу на базі інтерполяційних методів, кластеризації та районування, обробки великих даних, нейромережевого аналізу з використанням архітектури згорткових мереж на основі мови програмування Python в середовищі розробки ArcGIS ModelBuilder.

ВИСНОВКИ

В дисертаційній роботі на основі системного підходу розв'язано важливу науково-прикладну проблему інтелектуального аналізу геопросторових даних з метою розпізнавання прихованих закономірностей та відношень в задачах сталого розвитку територіально розподілених систем.

Результати дисертаційної роботи впроваджено в дослідженнях та звітах ННК «Світовий центр даних з геоінформатики та сталого розвитку» в межах дослідницьких робіт та ініціативних тематик аналізу параметрів сталого розвитку країн світу та України.

Результати дисертаційної роботи впроваджено в навчальний процес кафедри «Математичні методи системного аналізу» Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», зокрема при викладанні дисципліни «Геоінформаційні системи» та «Інтелектуальний аналіз в геоінформаційних системах».

Основні наукові та практичні результати роботи полягають у наступному:

1. Проведено аналіз дослідження проблем моделювання складових сталого розвитку та застосування методів геоінформатики для їх формалізації. За результатами цього аналізу зроблено висновок про доцільність розробки концепції класифікації та обґрунтування доцільності використання інтелектуальних методів аналізу геопросторових даних для моделювання сталого розвитку на засадах методів системного аналізу.

2. Системно обґрунтовано методи інтелектуального аналізу геопросторових даних в задачах сталого розвитку в межах концепції побудови вектору сталого розвитку на основі кватерного підходу до побудови компонентів якості життя та глобальних загроз. Запропоновано концепцію взаємодії між метрикою сталого розвитку, методами ІАГД та територією дослідження.

3. Розроблено методологію інтелектуального аналізу геопросторових даних в системах підтримки прийняття рішень для задач сталого розвитку з використанням теоретико-методологічних підходів до формалізації поняття та моделей

представлення геопросторових даних на основі парадигми дискретних та континуальних ознак тривимірного простору та його часової зміни. Основною метою використання інтелектуального аналізу геопросторових даних визначено пошук закономірностей та відносин у масивах даних, які містять просторово координовану прив'язку.

4. Досліджено застосування геоінформаційних технологій для проведення інтелектуального аналізу з використанням основних моделей організації геопросторових даних. Запропоновано систематизацію методів інтелектуального аналізу даних на основі геопросторових даних. Визначено ключові параметричні особливості цих методів, які полягають у геопросторовій координатності, топологічній єдності, геопросторових зв'язках та відносинах між собою. Визначено шість типів задач удосконалення метрики сталого розвитку, які можуть бути вирішені за допомогою ІАГД.

5. Розроблено систему методів, моделей та алгоритмів інтелектуального аналізу для обробки геопросторових даних для вирішення задач визначення впливу геопросторових відносин на складові сталого розвитку методом побудови трендових поверхонь. Застосовано методи геостатистичного аналізу даних для аналізу параметрів сталого розвитку України з використанням детермінованих та стохастичних методів аналізу даних, які включають процедури побудови глобальних інтерполяційних моделей на основі трендів, методу обернено зваженої відстані, сплайнів та методології крігінгу. В результаті доведено доцільність застосування трендових поверхонь для визначення глобальних трендів регіональної зміни параметрів сталого розвитку на прикладі регіонів України, що може бути застосовано у моделях передбачення та прогнозування.

6. Удосконалено методи геостатистичного аналізу та оптимізації задачі інтерполяції даних стохастичними методами крігінгу. Удосконалено застосування методу крігінгу при моделюванні невідомих значень на основі підбору сферичної моделі варіограми, що дозволило отримати найкращі статистичні оцінки точності для інтерпольованої поверхні методом сферичного крігінгу у порівнянні з детермінованими методами.

7. Запропоновано методи інтелектуального аналізу геопросторових даних на основі ГІС-технологій для аналізу безпекового компоненту на різних територіальних рівнях методами побудови просторово-часових кубів спільно з аналізом кластеризації «гарячих точок». Розроблено метод аналізу великих геопросторових даних шляхом їх дворівневого аналізу за допомогою організації даних у просторово-часових кубах, де на основі методів просторової кластеризації відбувається виділення інформаційних патернів даних. Апробація методики здійснена на базі даних про події у телекомунікаційній мережі оператора мобільного зв'язку, що дозволило виявити локальні тренди та кластери високих та низьких значень у розподілі подій.

8. Розроблено методи кластеризації регіонів України у багатовимірному просторі з врахуванням просторових відносин для проведення районування та зонування території. Розроблено методику районування території за параметрами сталого розвитку з використання топологічних геопросторових даних про місцезнаходження. Досліджено алгоритми деяких найбільш застосованих на практиці методів районування, які є модифікаціями канонічних методів, і метод бар'єрів максимальних відмінностей. Розроблено алгоритм побудови моделі сукупності однорідних територіальних зон. Оцінка ефективності угруповання об'єктів виконана за допомогою розрахунку псевдо-F-статистики Калінскі-Харабаза. Створені засоби проведення кластерного аналізу у середовищі ArcGIS дозволяють класифікувати просторові об'єкти за просторово-атрибутивними параметрами. Було здійснено апробацію розробленого підходу районування регіонів України на основі даних про параметри сталого розвитку для моделювання управлінських рішень з використанням методів просторової модифікації методу кластеризації к-середніх та нечітких к-середніх. Визначено, що застосування алгоритму нечітких к-середніх дає кращі статистичні результати.

9. Удосконалено методи глобальної та локальної просторової регресії для вирішення завдань оптимізації конфігурування просторових патернів у розподілі точкових об'єктів та методи геопросторової концептуалізації топологічних відносин між об'єктами. Розроблено метод моделювання на основі географічно

зваженої регресії та локального індексу Морана статистичного геопросторового розподілу індексу економічного виміру території.

10. Розроблено методи розрахунку метрики сталого розвитку на основі даних дистанційного зондування Землі з використанням методів карт самоорганізації Кохонена та згорткових нейронних мереж з градієнтним алгоритмом навчання.

Проведено навчання згорткової нейронної мережі на основі мультиспектральних зображень та знімків нічного освітлення, які дозволили з високою надійністю розпізнавати на знімках території міських та сільських населених пунктів, промислових зон та сільськогосподарських угідь. Це дозволило оцінити значення індексу економічного розвитку з використанням регресійних залежностей між типами землекористування та значеннями індексу економічного розвитку для окупованих територій України та АР Крим, сформувати розподіл значень індексу для адміністративних одиниць рівня району та міста, а також визначити динаміку індексу за останні роки.

В результаті можна зазначити, що використання методу згорткових нейронних мереж може дозволити отримати кількісні дані, що характеризують процес сталого розвитку території незалежно від статистичних джерел інформації з високим ступенем надійності. Основними властивостями цих даних є розрізнення вхідних космічних даних, крок мережі для апроксимації інформації та метод навчання нейронної мережі.

11. Розроблено загальну архітектуру інструментарію СППР на основі методології інтелектуального аналізу в задачах сталого розвитку, яка включає блоки постановки задач, управління даними, аналізу та моделювання, оцінювання результатів та прийняття рішень.

СППР реалізовано на основі веб-орієнтованої розподіленої геоінформаційної системи ArcGIS Enterprise 10.6. шляхом розробки аналітичних модулів геостатистичного аналізу на базі інтерполяційних методів, кластеризації та районування, обробки великих даних, нейромережевого аналізу з використанням архітектури згорткових мереж на основі мови програмування Python в середовищі розробки ArcGIS ModelBuilder.

12. Застосовано розроблену методологію та інструментарій до розв'язання ряду практичних задач моделювання параметрів сталого розвитку регіонів України, оцінки якості підземних вод, поводження з відходами, прогнозування локальних посух як фактору ризику для сільського господарства країни. На основі аналізу даних дистанційного зондування Землі проведено оцінку динаміки індексу економічного розвитку окремих міст України та окупованих територій. Визначено рівень зниження показників індексу економічного виміру для цих територій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Айвазян С.А. и др. Прикладная статистика: Исследование зависимостей. -М., Финансы и статистика, 1985. 607 с.
2. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д.^ Прикладная статистика: Классификация и снижение размерности. М., Финансы и статистика, 1989. - 607 с.
3. Айвазян С.А., Мхитарян В.С. Основы эконометрики. Теория вероятностей и прикладная статистика. М., Юнити, 2001. - 656 с.
4. Аналіз соціально-економічних процесів розвитку суспільства, заснованого на знаннях / наук. кер. проекту М. З. Згуровський. – К.: КПІ ім. Ігоря Сікорського, 2017. – 26 с. ISBN 978-966-622-819-5
5. Аналіз сталого розвитку – глобальний і регіональний контексти / Міжнар. рада з науки (ICSU) та ін.; наук. кер. проекту М. З. Згуровський. – К. : НТУУ «КПІ», 2013. – Ч. 2. Україна в індикаторах сталого розвитку (2013). – 178 с.
6. Арманд А.Д. Метод информационных градиентов в географическом районировании // Известия РАН. Серия географическая, №3 1973.
7. Берлянт А.М. Геоиконика. – М.: Fcnhtz, 1996. – 224 с.
8. Берлянт А.М. Геоинформатика: наука, технология, учебная дисциплина // Вестн. Моск. ун-та. Сер. геогр. – 1992. – № 2. – С. 1–23.
9. Берлянт А.М. Геоинформационное картографирование. – М.: Карт-геоцентр – Геодезиздат, 1997. – 64 с.
- 10.Бідюк П.І., Коршевніук Л.О. Проектування комп'ютерних інформаційних систем підтримки прийняття рішень: навч. посіб. / Навч.-наук. Комплекс «Ін-т прикл. систем. аналізу» НТУУ «КПІ», 2010. 340 с.
- 11.Блануца В.И. Интегральное экологическое районирование: концепции и методы. Новосибирск: ВО «Наука», 1993. - 159 с.
- 12.Блануца В.И. Преобразование метода районирования для решения задач географического прогнозирования. // География и природные ресурсы. -1987, №4.

- 13.Блануца В.И. Проблемный подход к районированию: построение алгоритма и опыт реализации. // География и природные ресурсы. 1989, №1.
- 14.Блануца В.И. Система методов автоматической классификации географических объектов: некоторые способы оценки качества классификаций // Известия РАН. Серия географическая, №3 1984.
- 15.Бочарников В.П. Fuzzy-технология: Математические основы. Практика моделирования в экономике. Санкт-Петербург: Наука РАН, 2001. 328 с.
- 16.Бродский Л. Проект INTAS по разработке автоматизированной технологии классификации земных покрытий: научные задачи, основные результаты и перспективы [Текст] / Л. Бродский, Е. И. Бушуев, В. И. Волошин, А. А. Козлова, О. И. Паршина, М. А. Попов, В. И. Саблина, А. И. Сахацкий, А. В. Сиротенко, Т. Соукуп, С. А. Станкевич, А. Г. Тарарико // Космічна наука і технологія. 2009. Т. 15. № 2. С. 36–48.
- 17.Визначення концентрації оксиду азоту в атмосферному повітрі за даними дистанційного зондування Землі / В.В. Путренко, Н.А. Куцина, С.Ю. Назаренко // Праці Центральної геофізичної обсерваторії. – К.: Інтерпрес ЛТД, 2017. – Вип. 13 (27) – С. 55 – 61.
- 18.Використання космічних знімків NPP/VIIRS у нічний час для оцінки економічної кризи на Сході України (Донецька та Луганська області). Лялько, В.І., Сахацький, О.І., Єлістратова, Л.О., Апостолов, О.А. Вісн. НАН України, 2017, № 2, С. 48-53.
- 19.Вистелиус А. Б. Математическая геология и ее вклад в фундаментальные геологические разработки. — Л., 1986. — 27 с. — (Препринт /ЛОМИ; Р-5-86).
- 20.Вистелиус А. Б. Математическая геология: история, состояние, перспективы. — Л., 1984. — 53 с. — (Препринт / ЛОМИ; Р-10-84).
- 21.Волкова В.Н. Теория систем и системного анализа. — М.: Наука, 1996.
- 22.Волошин А.Ф. Последовательный анализ вариантов в задачах исследования и проектирования сложных систем: монография / А.Ф.Волошин, В.И Кудин

- ; под общ. ред. И.В. Сергиенко. – К.: Издательско-полиграфический центр «Киевский университет», 2015. – 351 с.
23. Волошин О.Ф. Моделі та методи прийняття рішень: Навчальний посібник / О.Ф. Волошин, С.О. Мащенко. – К.: Видавничо- поліграфічний центр „Київський університет”, 2010. – 336 с.
24. Гайдышев И. Анализ и обработка данных: специальный справочник. СПб: Питер, 2001.-752 с.
25. Гандин Л. С., Каган Р. Л. Статистические методы интерполяции метеорологических данных. — Л.: Гидрометеиздат, 1976. — 359 с.
26. *Геоинформатика: учебник для студ. высш. учебн. заведений [Текст]. В 2 кн. Кн 1. / Е. Г. Капралов, А. В. Кошкарев, В. С. Тикунов и др. ; под ред. В. С. Тикунова. – 2-е изд. перераб. и доп. – М. : Издательский центр "Академия", 2008. – 384 с.*
27. Геоінформаційне картографування в Україні. Концептуальні основи і напрями розвитку. – За ред. акад. Л.Г. Руденка. – К., Науково-виробниче підприємство «Видавництво «Наукова думка» НАН України», 2011. – 103 с.
28. Геоінформаційне картографування техногенних ризиків на території України: формування баз даних // Вісник геодезії та картографії – 2010. – № 5. – С. 23 – 28
29. Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том 3: Просторові кадастрові інформаційні системи для інфраструктури просторових даних. Навчальний посібник. / М. Говоров, А.А. Лященко, Д. Кейк, П. Зандберген, М.А. Молочко, Л. Бевайніс, Л.М. Даценко, В.В. Путренко – Планета-Прінт, 2017. – 532 с.
30. Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том 2: Системи керування базами геоданих для інфраструктури просторових даних. Навчальний посібник. / Д. Кейк, А.А. Лященко, В.В. Путренко, Ю. Хмелевський, К.С. Дорошенко, М. Говоров – Київ, 2017. – 456 с.

31. Геоінформаційні технології та інфраструктура геопросторових даних: у шести томах. Том 1: Вступ до геоінформаційних систем для інфраструктури просторових даних. Навчальний посібник / Б. Магваїр, М. Говоров, Н.М. Пашинська, Л.М. Даценко, В.В. Путренко. – Київ, 2016. – 396 с.
32. Геоestatистика: теория и практика / В. В. Демьянов, Е. А. Савельева ; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. — М. : Наука, 2010. — 327 с.
33. Герасимов Б.М. Інтелектуальні системи підтримки прийняття рішень : навч. посіб. / Б. М. Герасимов, В. М. Локазюк, О.Г. Оксіюк, О. В. Поморова ; Європ. університет. – Київ, 2007. – 335 с.
34. Гермейер Ю.Б. Введение в теорию исследования операций. – М.: Наука, 1981.
35. Гусейн-Заде С.М., Тикунов В.С. Анаморфозы: что это такое? М.: Эдиторал УРСС, 1999.- 168 с.
36. Даценко Л.М., Говоров М., Путренко В.В., Пашинська Н.М. Перші результати реалізації українсько-канадського освітнього проекту з питань закладення основ інфраструктури просторових даних // Вісник геодезії та картографії, Вип.4, 2015. – С. 46-52
37. Дейт Л. Дж. Введение в системы баз данных. Пер. с англ. К.: Диалектика, 1998.
38. ДеМерс. Географические информационные системы. М: Дата+, 1999.-350 с.
39. Денисов А.А., Колесников Д.Н. Теория больших систем управления. – Л.: Энергоиздат, 1982.
40. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. Пер. с англ. - М.: Мир, 1984. - 333 с.
41. Довкілля України - 2013 [Текст]. – Державна служба статистики України, 2014. – 205 с.
42. Дрейпер Н., Смит Р. Прикладной регрессионный анализ // Н. Дрейпер, Р. Смит — М.: Статистика, 1973. — 138 с.
43. Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976.-512 с.

- 44.Дэйвинсон М. Многомерное шкалирование. Методы наглядного представления данных. М.: Финансы и статистика, 1988. - 348 с.
- 45.Дюк В., Самойленко А. Data Mining: учебный курс. СПб: Питер, 2001.-368 с.
- 46.Дюран Б., Оделл П. Кластерный анализ. М.: Финансы и статистика, 1977.- 176 с.
- 47.Евтеев, О. А. Проектирование и составление социально- экономических карт [Текст] : учебник / О. А. Евтеев. – М. : Изд-во Моск. ун-та, 1999. – 224 с.
- 48.Жамбю М. Иерархический кластер-анализ и соответствия. М: Финансы и статистика, 1989. -342 с.
- 49.Жуков В.Т., Сербенюк С.Н., Тикунов В.С. Математико-картографическое моделирование в географии. -М: Изд-во «Мысль», 1980. 223 с.
- 50.Журкин И.Г. Выбор критерия и пространства свойств для оценки эффективности инструментально-программных средств ГИС // Информационные технологии. – 1999. – № 3. – С. 28.
- 51.Журкин И.Г., Никишин А.Н. Анализ структуры данных для представления в ГИС // Геодезия и картография. – 2003. – № 8. – С. 44–49.
- 52.Заварзин А.В. Автоматическое районирование многомерных данных в векторных ГИС. – Дис. к.т.н., М., 2003. – 159 с.
- 53.Заварзин А.В., Тикунов В.С. Классификации // Основы геоинформатики. -М.: Изд-во «Академия», 2003. С. 213-248.
- 54.Заварзин А.В. Классификация многомерных данных при наличии ограничений // Сбор, матер, научно-практич. конф. ФАПСИ 12-13 ноября 2002 г. М: ГУИС ФАПСИ, 2002. - С. 188-193.
- 55.Заварзин А.В. Проблемы автоматического районирования больших объемов многомерных пространственных данных // Интернет Образование - Наука. Сбор, матер. III международ, конф. 8-12 октября 2002 г. - С.358-362.
- 56.Заварзин А.В., Месюра В.И. Визуализация автоматической классификации многомерных пространственных данных // Оптико-электронные и информационные технологии. 2002. №1(3). - С.30-38.

- 57.Заварзин А.В., Месюра В.И. Идентификация объектов при автоматической классификации многомерных пространственных данных в векторных ГИС // Измерительная и вычислительная техника в технологических процессах. - 2001. №4(18).-С.177-182.
- 58.Заварзин А.В., Месюра В.И. Разработка и исследование иерархических агломеративных алгоритмов районирования // Радиоэлектроника, информатика, управление. 2003. - №1 (9). - С.76-81.
- 59.Заварзин А.В., Орешкина Д.Д., Тикунов В.С. Электоральная культура России: классификации и картографирование в геоинформационной среде // Информационный бюллетень ГИС-Ассоциации. 2003. №1(38)-2(39). - С. 67-72.
- 60.Зайченко Ю.П. Исследование операций: учебник. Киев: «Слово», 2003. 688 с.
- 61.Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах / Ю.П. Зайченко – К.: «Издательский дом «Слово» », 2008. – 344 с.
- 62.Зайченко Ю.П. Теорія прийняття рішень: підручник / Ю.П. Зайченко – К.: НТУУ«КПІ», 2014.– 412 с.
- 63.Запровадження інтелектуальних інформаційних технологій для систем підтримки прийняття рішень / А. О. Морозов, В. Л. Косолапов, Л. В. Ромашкина, Р. П. Муренко // НТІ. – 2003. – № 3. – С. 36–41.
- 64.Зацерковний В.І. Геоінформаційні технології як засіб інтелектуалізації управління територіальними об'єктами // Вісник Астрономічної школи. – 2014. – Том 10, № 1-2. – 2014. – С. 97-101
- 65.Згуровский М., Панкратова Н. Системный анализ: Проблемы, методология, приложения. — К.: Наук. думка, 2005. — 743 с.
- 66.Згуровский М.З. Модели и методы принятия решений в нечетких условиях / М.З. Згуровский , Ю.П. Зайченко. – К.: НПП «Издательство “Наукова думка” НАН Украины», 2011. – 278 с.

- 67.Згуровский М.З. Основы вычислительного интеллекта / М.З. Згуровский , Ю.П. Зайченко. – К.: НПП «Издательство “Наукова думка” НАН Украины», 2013. – 408 с.
- 68.Згуровский М.З. Системный анализ. Проблемы, методология, приложения / М.З. Згуровский, Н.Д. Понкратова. – 2-е изд., перераб. и доп. – К.: Наукова думка, 2011. – 725 с.
- 69.Згуровский М.З., Павлов О.А. Принятие решений в сетевых системах с ограниченными ресурсами. Киев: Наукова думка, 2010.
- 70.Згуровский М.З., Панкратова Н.Д. Технологическое предвидение. Киев: ІВЦ "Видавництво "Політехніка", 2005. 155 с.
- 71.Згуровський М.З. Інтелектуальний аналіз і системне погодження наукових даних в міждисциплінарних дослідженнях / Згуровський М.З., А.О. Болдак, К.В. Єфремов // Кибернетика и системный анализ. – 2013. – № 4. – С. 62-75.
- 72.Иванов В.И., Легостаев С.Е. Цифровые методы картографирования. - Геодезия и картография. 1989. №10.
- 73.Іщук О.О. Просторовий аналіз і моделювання в ГІС: Навч. посібник / О. О. Іщук, М. М. Коржнев, О. Е. Кошляков; за ред. акад. Д. М. Гродзинського. – К. : Вид.-поліграфічний центр "Київський університет", 2003. – 200 с.
- 74.Калянов Н.Г. CASE структурный системный анализ (автоматизация и применение). -М: «ЛОРИ», 1996.
- 75.Каневский М., Демьянов В., Савельева Е. и др. Элементарное введение в геостатистику. — М., 1999а. — 136 с. — (Проблемы окружающей среды и природных ресурсов / ВИНТИ; № 11).
- 76.Каневский М., Демьянов В., Чернов С. и др. Геостатистика и искусственные нейронные сети для анализа и моделирования пространственно распределенных данных // Изв. РАН. Энергетика. — 1999б. — № 1.
- 77.Каплан Р., Фалб М., Арбиб М. Очерки по математической теории систем. — М.: Мир, 1971.

- 78.Карпик А.П. Концепция геоинформационного пространства // Материалы международной научно-технической конференции, посвященной 225-летию МИИГАиК. Ч. Геодезия. – М.: МИИГАиК, 2004. – С. 434–438.
- 79.Карпик А.П. Методологические и технологические основы геоинформационного обеспечения территорий: монография. – Новосибирск: СГГА, 2004.
- 80.Карпик А.П. Управление территорией в геоинформационном дискурсе [Текст]: монография / А.П. Карпик, А.Г. Осипов, П.П. Мурзинцев. – Новосибирск: СГГА, 2010. – 280 с.
- 81.Кини Р. Л., Райфа Х. Принятие решений при многих критериях предпочтения и замещения [пер. с англ. под ред. Шахнова И. Ф.] / Р.Л. Кини, Х. Райфа. – М.: Радио и связь, 1981. – 560 с.
- 82.Кисельова Я.Г. Інтелектуальна підтримка прийняття рішень по забезпеченню безпеки об'єктів в умовах надзвичайних ситуацій / Я. Г. Кисельова // Системи озброєння і військ. техніка. – 2008. – № 4. – С. 64–66.
- 83.Кісь Я.П. Інтелектуальні геоінформаційні системи. Міжнародний досвід та шляхи розвитку в Україні / Я. П. Кісь, Н. Б. Шаховська, О. Б. Вальчук // Вісн. Нац. ун-ту “Львів. політехніка”. Сер. Інформ. системи та мережі. – 2008. – № 621. – С. 139–144.
- 84.Классификация и кластер. М.: Мир, 1980. - 392 с.
- 85.Комп'ютерна програма «Геоінформаційна система «Гідровузли України»:свідoctво про реєстрацію авторського права на твір 64745 Україна / Бенатов Д. Е.,Путренко В. В., Стефанишин Д. В., Розщупкін Г. В. – № 65227; заявл. 09.02.16; реєстр. 01.04.16.
- 86.Коновалова Н.В., Капралов Е.Г. Введение в ГИС. Петрозаводск: изд-во Петрозавод. ун-та, 1995. - 148 с.
- 87.Королев Ю.К. Общая геоинформатика. М: Дата+, 1998. - 118 с.
- 88.Королев Ю.К. Статьи, лекции, доклады по проблемам геоинформатики. М: Дата+, 2000. - 127 с.

- 89.Кохан С. С. Застосування вегетаційних індексів на основі серії космічних знімків IRS-1D LISS-III для визначення стану посівів сільськогосподарських культур / С.С. Кохан // Космічна наука і технологія. - 2011. - Т. 17. - № 5. - С. 58–63.
- 90.Кочергин Г.А. Моделирование однородных территориальных зон на основе многомерной кластеризации и ГИС-анализа в условиях малого объема данных. – Дис. к.т.н., Ханты-Мансийск, 2011. – 138 с.
- 91.Кретов В.С., Пинчук И.С., Заварзин А.В. Использование геоинформационных систем при планировании и проведении миротворческих операций // Военная мысль. 2001, №6. - С.23-27.
- 92.Ларичев О.И., Мошкович Е.М. Качественные методы принятия решений. Москва: Наука, 1996. 208 с. Згуровский М.З., Панкратова Н.Д. Системный анализ: проблемы, методология, приложения. Киев: Наукова думка, 2005. 743с.
- 93.Левинский С.Е. Таксономические методы в региональных исследованиях. - В сб.: Региональная наука о размещении производительных сил. Новосибирск-Иркутск, 1971.
- 94.Лисицкий Д.В. Общность и различие понятий «цифровая модель местности», «цифровая карта» и «электронная карта» // Современные проблемы геодезии и оптики. LI научно-техн. конф., 16–19 апр. 2001 г.: тез. докл. – Новосибирск: СГГА, 2001. – С. 143–144.
- 95.Лялько В.І. Особливості прогнозування врожайності зернових культур забагато спектральними даними ДЗЗ / В.І. Лялько, О.І. Сахацький, Г.М. Жолобак // Багатоспектральні методи дистанційного зондування Землі в задачах природокористування / За ред. В. І. Лялька, М. О. Попова. — Київ : Наук. думка, 2006. — С. 176—191.
- 96.Матерон Ж. Основы прикладной геостатистики. — М.: Мир, 1968. —407 с.
- 97.Мозговой Д.К. Использование многоспектральных снимков для классификации посевов сельхозкультур / Д.К. Мозговой, О.В.Кравец // Екологія та ноосферологія. - 2009. - Т. 20. - № 1–2. – С. 54-58

98. Мурзинцев П.П., Чахлова А.П. Инженерное управление территориями (Использование ГИС MapInfo в зонировании городов): учеб.-метод. пособие. – Новосибирск: СГГА, 2008. – 106 с.
99. Мыльник В.В. Инвестиционный менеджмент. – М.: ИНФО, 2000.
100. Нутенко Л.Я. Математические методы регионализации (некоторые вопросы теории и практического применения). Автореферат диссертации на соискание ученой степени кандидата географических наук. М., 1970.-30 с.
101. Нутенко Л.Я. Типизация географических объектов как задача теории распознавания образов. // Известия РАН. Серия географическая, №6 1971.
102. Опыт создания ГИС-проектов в среде программы WinGIS / В.А. Середович, А.П. Карпик, О.В. Твердовский, С.В. Середович // Тез. докл. на междунаrod. научно-техн. конф., посвящ. 65-летию СГГА «Соврем. проблемы геодезии и оптики». – Новосибирск: СГГА, 1998. – С. 113.
103. Осипов А.Г. Управление территорией. Актуальные проблемы: монография / Под общ. ред. И.В. Лесных. – Новосибирск: СГГА, 2003. – 378 с.
104. Панкратова Н.Д., Савченко І.О. Морфологічний аналіз. Проблеми, теорія, застосування/ навч. посібник. Київ: Наукова думка. 2015. 245 с.
105. Пархоменко А.В., Путренко В.В. Аналіз регіональних збройних конфліктів з використанням методів просторової кластеризації // Системний аналіз та інформаційні технології: матеріали 19-ї Міжнародної науково-технічної конференції SAIT 2017, Київ, 22 – 25 травня 2017 р. / ННК “ІПСА” НТУУ “КПІ ім. Ігоря Сікорського”. – К.: ННК “ІПСА” НТУУ “КПІ”, 2017. – С. 101-102.
106. Пашинська Н.М., Путренко В.В. Використання інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки // Геоінформаційні технології у територіальному управлінні – Одеса : ОРІДУ НАДУ, 2016. – 123-126
107. Пашинська Н.М., Путренко В.В. Використання карт самоорганізації для кластеризації території за рівнем природно-техногенної небезпеки //

- Міждисциплінарні дослідження актуальних проблем застосування інформаційних технологій в сучасному світі: зб. матеріалів V Всеукр. наук.-практ. конф. "Глушковські читання", Київ, 24 листопада 2016 року. – К: ТОВ НВП "Інтерсервіс", 2016. – С. 145-150.
108. Пашинська Н.М., Путренко В.В. Інтелектуальний аналіз даних на основі класифікації факторів пожежної небезпеки з використанням дерева рішень // Праці VIII міжнародної школи-семінару «Теорія прийняття рішень». – Ужгород, УжНУ, 2016. - С. 203-204
 109. Пашинська Н.М., Путренко В.В. Просторово-часове моделювання забруднення повітря внаслідок аварій на хімічних об'єктах з використанням інструментарію ГІС // Системний аналіз та інформаційні технології: матеріали 19-ї Міжнародної науково-технічної конференції SAIT 2017, Київ, 22 – 25 травня 2017 р. / ННК "ІПСА" НТУУ "КПІ ім. Ігоря Сікорського". – К.: ННК "ІПСА" НТУУ "КПІ", 2017. – С. 103-104
 110. Перспективи розвитку Національного атласу України в мультимедійному напрямі // Вісник геодезії та картографії – 2009. – № 6. – С. 14 – 20.
 111. Петренко А.І. Grid та інтелектуальна обробка даних Data Mining / А.І. Петренко // Систем. дослідж. та інформ. технології. — 2008. — № 4. — С. 97-110.
 112. Питенко А.А. Нейросетевой анализ в ГИС. // Автореф. дисс. канд. техн. наук 05.13.16, Красноярск, 2000.
 113. Попов М. А. Использование полного набора нормализованных межканальных индексов многоспектральных космических изображений при классификации покрытий ландшафта [Текст] / М. А. Попов, С. А. Станкевич, А. И. Сахацкий, А. А. Козлова // Уч. зап. Таврического нац. ун-та им. В. И. Вернадского. — 2007. — 20 (59), № 1. — С. 175–182.
 114. Путренко В. В. Використання хмарних сервісів для тематичного картографування // Ученые записки Таврического национального

- університета ім. В.І. Вернадського, Серія «Географія» 2012. – Том 25 (64) №1. – С. 191 – 200.
115. Путренко В. В. Геоинформационная оценка рисков и опасностей техногенных чрезвычайных ситуаций (на примере Винницкой области) // ИнтерКарто-ИнтерГИС – 18: Устойчивое развитие территорий: теория ГИС и практический опыт. Материалы международной конференции / Редкол.: С.П. Евдокимов (отв. ред.) [и др.]. Смоленск, 26-28 июня, 2012 г. Смоленск, 2012. – С. 367- 374.
 116. Путренко В. В. Засади створення регіональної ГІС «Ресурсозбереження» // Національне картографування: стан, проблеми та перспективи розвитку: Збірник наукових праць / Відп. За вип. А.А. Москалюк. – К.: ДНВП «Картографія», 2008. – Вип.3 – С. 207 – 212.
 117. Путренко В. В. Картографічне моделювання геохімічних полів // Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія: Географічні науки. - 2011 – Т25, №3. – С. 144 – 152
 118. Путренко В. В. Картографічний метод оцінки екологічної безпеки використання підземних вод в Україні // Вчені записки Таврійського національного університету ім. В.І. Вернадського. Серія: Географічні науки. - 2011 –Т 24., №2. С.54-62
 119. Путренко В. В. Картографічний підхід до оцінювання техногенних ризиків виникнення надзвичайних ситуацій на регіональному рівні // Вісник геодезії та картографії. – 2012. - №6. – С. 24-32.
 120. Путренко В. В. Картографування інфраструктурних мереж для цілей управління ризиками виникнення надзвичайних ситуацій // Проблеми безперервної географічної освіти і картографії: Збірник наукових праць. – Харків: ХНУ ім. В.Н. Каразіна, 2012. – Вип. 16. – С. 106 – 109.
 121. Путренко В. В. Підходи до класифікації географічних баз даних // Проблеми безперервної географічної освіти і картографії: Збірник наукових праць. – Харків: ХНУ ім. В.Н. Каразіна, 2011. – Вип.13. – С. 78-81

122. Путренко В. В., Чабанюк В.С., Станкевич Т.В. Питання публікації тематичної геопросторової інформації на основі картографічних веб-сервісів // Український географічний журнал, 2012. - № 4. – С. 42 – 52.
123. Путренко В. В., Голубцов О.Г., Чехній В.М., Фаріон Ю.М. Ландшафтна ГІС як результат ландшафтознавчого прикладного дослідження адміністративного району: методичні аспекти // Географія та туризм, 2010. – Вип. 10. – С. 141 – 153
124. Путренко В. В., Ляшенко Д.О. Математико-картографічне моделювання електоральної ситуації в Україні // Часопис картографії: Збірник наукових праць. – К.: КНУ ім. Тараса Шевченка, 2011. – Вип.2. – С. 86-95
125. Путренко В. В., Мультимасштабні картографічні моделі аналізу небезпек та ризиків надзвичайних ситуацій // Ученые записки Таврического национального университета имени В.И.Вернадского Серия «География». Том 26 (65). 2013 г. № 1, С. 136-144.
126. Путренко В. В., Пашинська Н. М., Назаренко С. Ю. Картографування якості повітря на основі аналізу даних дистанційного зондування // Часопис картографії: Збірник наукових праць. – К. : КНУ ім. Тараса Шевченка, 2016. – Вип. 15. Частина 1. – 89-98
127. Путренко В. В., Півньов О.В Створення та функціонування бази даних Національного атласу України // Укр. геогр. журн. – 2010. – № 1. – С. 53-58.
128. Путренко В. В., Тихоход В.А. Використання кластерного аналізу для просторового моделювання в геоінформаційних системах // Географічна наука і практика: виклики епохи: Матеріали міжнародної наукової конференції, присвяченої 130-річчю географії у Львівському університеті (м. Львів, 16 – 18 травня 2013 р.). У 3-ох томах. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2013. – Том 3. – 306 с.
129. Путренко В.В Системные основы интеллектуального анализа геопространственных данных / В.В. Путренко // Системні дослідження та інформаційні технології. – 2015. – № 3 - С. 24 – 32.

130. Путренко В.В. База даних ландшафтознавчого дослідження адміністративного району з метою моніторингу та раціонального використання земель
131. Путренко В.В. Базы данных тематического картографирования в составе инфраструктур пространственных данных // Тематическое картографирование для создания инфраструктур пространственных данных / Материалы IX научной конференции по тематической картографии (Иркутск, 9-12 ноября 2010 г.). – Иркутск: Изд-во Института географии им. В.Б. Сочавы СО РАН, 2010. – В 2-х т. – Т. 1. – С. 56 – 59.
132. Путренко В.В. Використання глобальних моделей інтерполяції у інтелектуальному аналізі геопросторових даних / В.В. Путренко // Системний аналіз та інформаційні технології: матеріали 17-ї Міжнародної науково-технічної конференції SAIT 2015, Київ, 22-25 червня 2015 р. / ННК “ІПСА” НТУУ “КПІ”. – К.: ННК “ІПСА” НТУУ “КПІ”, 2015. – С. 103 – 104.
133. Путренко В.В. Геоінформаційне атласне картографування природокористування // Українська історична географія та історія географії в Україні: Матеріали міжнародної наукової конференції (7-10 жовтня). Чернівці: Чернівецький національний університет, – 2009. – С. 176.
134. Путренко В.В. Геоінформаційний аналіз регіональної інфраструктури з метою попередження надзвичайних ситуацій // Геопросторові технології в сталому розвитку міст: матеріали Міжвузівської науково-практичної конференції, Харків, - 16- 17 листопада 2011 р. – Х.: ХНАМГ, 2011. – С. 16-18
135. Путренко В.В. Екологічний моніторинг засобами дистанційного зондування Землі // ЕКОЛОГІЯ ЛЮДИНИ – РОДИНИ – КРАЇНИ: зб. матеріалів круглого столу, (22 листоп. 2018 р., м. Київ) – Київ : КПІ ім. Ігоря Сікорського, 2018. – С. 62 – 64.
136. Путренко В.В. Засади сервіс-орієнтованого картографування для попередження надзвичайних ситуацій // Матеріали Всеукраїнської

- конференції «На шляху до сталого розвитку», 18 – 19 червня 2013 р., Київ, с. 120 – 124.
137. Путренко В.В. Застосування трендових геопросторових статистичних моделей на основі інтелектуальному аналізі даних/ В.В. Путренко // Системні дослідження та інформаційні технології. – 2015. – № 4. – С7 – 19.
 138. Путренко В.В. Інтелектуальний аналіз геопросторових даних: сучасні тенденції та перспективи // Международный научно-практический форум «НАУКА И БИЗНЕС», 1 июля 2016 года Днепр, Украина. – Днепр, 2016. – С. 76 – 82.
 139. Путренко В.В. Інтелектуальний аналіз геопросторових даних для цілей управлінської діяльності / В.В. Путренко // Матеріали Всеукраїнської науково-практичної конференції «В.М. Глушков – піонер кібернетики» (2014 р. м. Київ). – К.: Видавництво «Політехніка», 2014. – С.235 – 237.
 140. Путренко В.В. Інтелектуальний аналіз земного покриття території України на основі даних Globeland 30 // Вісник Астрономічної школи, 2016, том 12, № 1 – С. 46 – 57.
 141. Путренко В.В. Інтелектуальний аналіз небезпеки виникнення природних пожеж на основі геоінформаційних технологій // Технологический аудит и резервы производства. – Х., Технологический центр. – 2016. – С. 67 – 72.
 142. Путренко В.В. Методи інтелектуального аналізу геопросторових даних // Матеріали XII Міжнародної науково-технічної конференції «Авіа 2015» 28 – 29 квітня 2015 року, –К.: НАУ, 2015. – С. 28.43 – 28.48.
 143. Путренко В.В. Мультимасштабні геопросторові бази даних картографування ризиків та небезпек надзвичайних ситуацій // «Потенціал сучасної географії у розв’язанні проблем розвитку регіонів», Київ 3-5 жовтня 2013 р., с. 432 -435.
 144. Путренко В.В. Нові технологічні засади екологічного картографування // Зелена економіка: перспективи впровадження в Україні: матеріали

- міжнародної конференції (Київ, 24 – 25 квітня 2012 р.): у 3 т. – К. : Центр екологічної освіти та інформації, 2012. – Т.1. – С. 369 – 372.
145. Путренко В.В. Пашинська Н.М. “Big Data” в просторовому управлінні міським середовищем // Геоінформаційні технології у територіальному управлінні. – Одеса: ОРІДУ, 2015. – С. 89-93
 146. Путренко В.В. Підходи до формування бази даних атласу надзвичайних ситуацій // Просторовий аналіз природних і техногенних ризиків в Україні: Зб. Наук. праць. – 2009. – С. 37 – 40.
 147. Путренко В.В. Побудова геоінформаційних баз даних управління природними ресурсами // Природно-ресурсний потенціал збалансованого (сталого) розвитку України: Матеріали Міжнародної науково-практичної конференції (Київ, 19-20 квітня 2011 р.) у 2 т. – К.: Центр екологічної освіти та інформації, 2011. – Т.1. – С. 424 – 427
 148. Путренко В.В. Ресурсозбереження: регіональний аспект. - Наук. ред. Лісовський С.А. – К.: Інститут географії НАН України, 2011. – 224 с.
 149. Путренко В.В. Світовий досвід організації тематичної інформації у інфраструктурах геопросторових даних // Розвиток тематичної складової інфраструктури геопросторових даних в Україні: Збірник наукових праць. – К., 2011. – С. 133 – 139
 150. Путренко В.В. Статистичний аналіз даних землекористування для цілей регіонального управління в Україні // Міжнародна наукова конференція Астрономічна школа молодих вчених Україна, Київ, 26–27 травня 2016 р. – К., НАУ. – 2016. – С. 77 – 79.
 151. Путренко В.В., Бенатов Д. Е., Стефанишин Д . В . Геоінформаційна система «гідровузли України» – важливий елемент підтримки управлінських процедур // Восточно-Европейский журнал передовых технологий. - 1/3 (79) 2016. – 46 – 53.
 152. Путренко В.В., Джигирей І.М. ГІС-технології як ефективний інструмент залучення громадськості у процедури ухвалення рішень з питань екологічної політики / В.В. Путренко, І.М. Джигирей // Збірка тез

- доповідей XVII Міжнародної науково-практичної конференції «Екологія. Людина. Суспільство» (м. Київ) / Укладач Д.Е. Бенатов. – К.: НТУУ «КПІ», 2015. – С. 183-185.
153. Путренко В.В., Красовська І.Г. Кластеризація геопросторових даних при інтелектуальному аналізі // Радіоелектронні і комп'ютерні системи. – ХАІ. – 2015. - № 3. – С. 32-41.
 154. Путренко В.В., Назаренко С.Ю. Визначення якості повітря на основі інтелектуального аналізу даних дистанційного зондування // Математичне моделювання в економіці, 2016. – № 3-4 (7). – С. 176-188.
 155. Путренко В.В., Назаренко С.Ю., Пашинська Н.М. Моделювання даних з просторово-часовим розподілом для потреб прийняття рішень // Містобудування та територіальне планування: Наук.-техн. збірник / Головн. ред. М.М. Осетрін. – К., КНУБА, 2018. – Вип. 68. – С. 750 – 760.
 156. Путренко В.В., Пархоменко А.В. Геопросторовий аналіз та візуалізація даних про військову конфліктогенність // Містобудування та територіальне планування: Наук.-техн. збірник / Головн. ред. М.М. Осетрін. – К., КНУБА, 2017. – Вип. 64. – С. 326 – 332.
 157. Путренко В.В., Пашинська Н.М. Використання 3D Space-Time Cube для інтелектуального аналізу даних // System analysis and information technology: 20-th International conference SAIT 2018, Kyiv, Ukraine, May 21- 24, 2018. Proceedings. – pp. 150-151
 158. Путренко В.В., Пашинська Н.М. Використання 3D Space-Time Cube для моделювання просторово-часових даних // Збірник матеріалів конференції «ГІС-ФОРУМ-2018» (Харків, 14–16 березня 2018 р.). — Вип. 2. — Х.: ХНУ імені В. Н. Каразіна. – С.32-34.
 159. Путренко В.В., Пашинська Н.М. Інтелектуальний аналіз геопросторових даних як інструмент планування сталого розвитку // Вісник Київського національного університету імені Тараса Шевченка. Серія фізико-математичні науки. 2017, Вип 1. С. 49-55

160. Путренко В.В., Пашинська Н.М. Інтелектуальний аналіз земного покриття території України на основі даних Globeland 30 // Управління розвитком складних систем. - Вип 31. 2017. – С. 117-126
161. Путренко В.В., Пашинська Н.М. Оцінка параметрів сталого розвитку методами ДЗЗ // Картографія та вища школа сучасний стан і стратегія розвитку. Матеріали міжнародної конференції. – К., 2018. – С. 71-74
162. Путренко В.В., Пашинська Н.М. Регіональний моніторинг стану рослинності з використанням методів інтелектуального аналізу даних // Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки. - Том 30 (69) Ч. 1 № 4 2019. – С. 112 – 118.
163. Путренко В.В., Руденко Л.Г. Атлас ризиків та небезпек виникнення надзвичайних ситуацій // Національне картографування: стан, проблеми та перспективи розвитку, Збірник матеріалів V Всеукраїнської науково-практичної конференції «Національні атласи у формуванні глобального інформаційного простору» (м. Київ, 13-14 вересня 2012 р.), Вип. 5. – 2012. – С. 155 – 159.
164. Путренко В.В., Руденко Л.Г., Волошин С.М. Інформаційний продукт щодо аналізу ризиків та небезпек виникнення надзвичайних ситуацій в Україні // Екологічні аспекти регіонального партнерства в надзвичайних ситуаціях: Збірка матеріалів I Міжвузівської науково-методичної конференції. – Х.: НУЦЗУ, 2012. – С. 97 – 103.
165. Путренко В.В., Тихоход В.О. Розробка геостатистичних моделей інтелектуального аналізу забруднення повітря // Східно-Європейський журнал передових технологій. – 2015. - №4/10. – С. 21-26.
166. Путренко В.В., Чабанюк В.С., Дишлик О.П. Головні технологічні рішення при створенні національних атласів // Національне картографування: стан, проблеми та перспективи розвитку, Збірник матеріалів V Всеукраїнської науково-практичної конференції «Національні атласи у формуванні глобального інформаційного простору» (м. Київ, 13-14 вересня 2012 р.), Вип. 5. – 2012. – С. 177 – 180

167. Путренко, В. В. Застосування інструментарію геоінформаційного моделювання для інтелектуального аналізу даних пожежної небезпеки / В. В. Путренко, Н. М. Пашинська // *Вісник НТУ «ХПІ», Серія: Нові рішення в сучасних технологіях*. – Харків: НТУ «ХПІ». – 2017. – 7 (1229). – С. 156-163. – doi:10.20998/2413-4295.2017.07.22.
168. Саати Т.Л. Принятие решений при зависимостях и обратных связях: Аналитические сети. Москва: Книжный дом «ЛИБРОКОМ», 2009. 360 с.
169. Сарычева Л.В. Пространственно-временной подход в задачах кластеризации / Л.В. Сарычева // *Штучний інтелект*. – 2006. – № 3. – С. 646-653.
170. Світличний, О. О. Основи геоінформатики [Текст] : навчальний посібник / О. О. Світличний, С. В. Злотницький ; за заг. ред. О. О. Світличного. – Суми : ВТД Університетська книга, 2006. – 295 с.
171. Сербенюк, С. Н. Картография и геоинформатика – их взаимодействие [Текст] / С. Н. Сербенюк. – М., 1990. – 159 с.
172. Сердюцкая Л. Ф., Яцишин А. В.- Техногенная экология. Математико-картографическое моделирование. — М.: Книжный дом "ЛИБРОКОМ", 2009.-232с.
173. Ситник В.Ф. Інтелектуальний аналіз даних (дейтамайнінг) [Текст]: навч. посібник / В. Ф. Ситник, М.Т. Краснюк. – К. : КНЕУ, 2007. – 376 с.
174. Станкевич С. Комплексна обробка багатоспектральних аерокосмічних зображень для виявлення сільськогосподарських угідь / С. Станкевич, А. Васько // *Сучасні досягнення геодезичної науки та виробництва : збірник наукових праць Західного геодезичного товариства УТГК / Західне геодезичне товариство Українського товариства геодезії і картографії, Національний університет "Львівська політехніка" ; головний редактор І. С. Тревого*. – Львів : Видавництво Львівської політехніки, 2011. – Випуск 2 (22). – С. 171-175.
175. Тикунов В.С. Алгоритм для моделирования тематического содержания типологических карт. Вестник Моск. ун-та, сер. геогр., 1983, N 4, с. 78-84.

176. Тикунов В.С. Атласная информационная система «Устойчивое развитие России» // Вестник МГУ, сер. Геогр. 2002, № 5. - С. 21-32.
177. Тикунов В.С. Классификации в географии: ренессанс или увядание? - Смоленск, Изд-во СГУ, 1997 367 с.
178. Тикунов В.С. Моделирование в картографии. М: Изд-во МГУ, 1997, 405 с.
179. Тикунов В. С. Геоинформатика: Учеб. для студ. вузов / Е.Г.Капралов, Г35 А. В. Кошкарёв, В. С. Тикунов и др.; Под ред. В. С.Тикунова. — М : Издательский центр «Академия», 2005. — 480 с.
180. *Тикунов, В. С. Классификации в географии: ренессанс или увядание? [Текст] / В. С. Тикунов.— М.,-Смоленск : Изд-во СГУ, 1997. – 367 с.*
181. Трофимов А.М., Заботин Я.И., Панасюк М.В., Рубцов В.А. Количественные методы районирования и классификации. Казань: изд-во Казанск. ун-та, 1985.- 120 с.
182. Ту Дж., Гонсалес Р. Принципы распознавания образов. М.: Мир, 1978.
183. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. М.: ИНФРА-М, 1998. - 528 с.
184. Уткин Л.В. Анализ риска и принятие решений при неполной информации. Санкт-Петербург: Наука, 2007. 404 с.
185. Форсайт економіки України: середньостроковий (2015–2020 роки) і довгостроковий (2020–2030 роки) часові горизонти (версія для обговорення) / наук. ке- рівник проекту акад. НАН України М. З. Згуровський // Міжнародна рада з науки (ICSU); Національний технічний університет України «Київський політехнічний інсти- тут»; Інститут прикладного системного аналізу НАН України і МОН України; Світовий центр даних з геоінформатики та сталого розвитку. – Київ : НТУУ «КПІ», 2015. – 36 с.
186. Форсайт та побудова стратегії соціально-економічного розвитку України на середньостроковому (до 2020 року) і довгостроковому (до 2030 року) часових горизонтах / наук. керівник проекту акад. НАН України М. З. Згуровський // Міжнародна рада з науки (ICSU); Комітет із системного

- аналізу при Президії НАН України; Національний технічний університет України «Київський політехнічний інститут»; Інститут прикладного системного аналізу МОН України і НАН України; Світовий центр даних з геоінформатики та сталого розвитку. — Київ : НТУУ «КПІ», 2016. — 180 с.
187. Хурцилава К. В. К вопросу о пространственно-временных ГИС и некоторых их применениях / Хурцилава К. В., Білецький Б.О., Беспалов В.П. // Системи підтримки прийняття рішень. Теорія і практика. — Київ: ПІММС НАНУ. — 2013. - С. 52-55.
 188. Цвиркун А.Д. Структура сложных систем. — М.: Машиностроение, 1975.
 189. Циганок В.В. Удосконалення методу цільового динамічного оцінювання альтернатив та особливості його застосування. Реєстрація, зберігання і обробка даних. 2013. Vol. 15, No 1. Р. 90–99.
 190. Черняк О.І. Інтелектуальний аналіз даних: Підручник / О.І. Черняк, П.В. Захарченко; Київ. нац. ун-т ім. Т. Шевченка. — К. : Знання, 2014. — 599 с.
 191. Щокін В.П. Інтелектуальні системи керування: аналітичний синтез та методи дослідження / В.П. Щокін. — Кривий Ріг : Д.О. Чернявський, 2010. — 264 с.
 192. Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. In: Proceedings of international conference on very large databases (VLDB), Santiago, Chile, pp 487–499
 193. Aji A et al (2013) Hadoop-GIS: a high performance spatial data warehousing system over MapReduce. In: Proceedings of the 39th international conference on very large data bases, VLDB endowment, vol 6(11), pp 1009–1020 (August 26–30th 2013, Riva del Garda, Trento, Italy)
 194. Al G (1998) The digital earth: understanding our planet in the 21st century, speech at the California science center, Los Angeles, California, on January 31, 1998. url:http://www.isde5.org/al_gore_speech.htm

195. Ambroise C., Govaert G. Spatial Clustering and the EM Algorithm. France: Universite de technologie de Compienge, 1996.
196. Andrew B. Lawson. Bayesian disease mapping. Hierarchical modeling in spatial epideviology. Chapman and Hall/CRC, 2009. - 344p.
197. Andrienko G. Data mining with C4.5 and interactive cartographic visualization / G. Andrienko, N. Andrienko // In N. W. G. T. Paton (Ed.), User interfaces to data intensive systems. Los Alamitos, CA: IEEE Computer Society. – 1999. – pp. 162–165.
198. Andrienko N. Exploratory SpatioTemporal Visualization: An Analytical Review / N. Andrienko, G. Andrienko, P. Gatalsky // J. Visual Languages and Computing. - vol. 14. - no. 6, 2003. - pp. 503-541.
199. Andrienko, N., Andrienko, G., Gatalsky, P.: SpatioTemporal Visualization: An Analytical Review J. Visual Languages and Computing, vol. 14, no. 6, (2003) 503-541
200. Armstrong M. Basic Linear Geostatistics. — [S. l.]: Springer Verl., 1997.
201. Arora V., Lieskovsky J. Electricity Use as an Indicator of U.S. Economic Activity. (Washington, U.S. Energy Information Administration, 2014). http://www.eia.gov/workingpapers/pdf/electricity_indicator.pdf.
202. Arthurs AM (1965) Probability theory. Dover Publications, London
203. *Ashkan Zarnani, Masoud Rahgozar* Mining Spatial Trends by a Colony of Cooperative Ant Agents // Spatial Data Mining: Consolidation and Renewed Bearing, April 22, 2006. P.64 – 72.
204. Barnett V (1978) Outliers in statistical data. Wiley, New York
205. Barrabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512
206. Beckschäfer P. Mapping leaf area index in subtropical upland ecosystems using rapideye imagery and the randomforest algorithm / P. Beckschäfer, L. Fehrmann, R.D. Harrison, J. Xu, C. Kleinn // iForest, - 2014. - 7. - P. 1–11
207. *Blackmore B. S.* (2000). The interpretation of trends from multiple yield maps. Computers and Electronics in Agriculture, 26(1), 37–51

208. *Blackmore1 Simon; Godwin Richard J.; Fountas Spyros* The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years Biosystems Engineering (2003) 84 (4), P. 455–466
209. Boken V.K. Improving an operational wheat yield model using phenological phase-based Normalized Difference Vegetation Index / V.K. Boken, C.F. Shaykewich // International Journal of Remote Sensing. – 2002. - 23. - P. 4155–4168
210. Bolton D.K. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics / D.K. Bolton, M.A. Friedl // Agric. For. Meteorol. – 2013. - 173. - P. 74–84
211. Bossard M., Feranec J., Otahel J. The revised and supplemented Corine Land Cover nomenclature // Techn. Rept EEA. — 2000. — N 38. — 110 p.
212. Brian O'Connor, Cristina Secades, Johannes Penner, Ruth Sonnenschein, Andrew Skidmore, Neil D. Burgess & Jon M. Hutton Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets // Remote Sensing in Ecology and Conservation/ - y John Wiley & Sons Ltd, 2015/ - 19 – 27
213. Broge N.H. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density / N.H. Broge, E. Leblanc // Remote Sens. Environ. – 2001. – 76. - P. 156–172
214. Buckless BP, Petry FE (1994) Genetic algorithms. IEEE Computer Press, Los Alamitos
215. *Burrough P.A., McDonnel, R.A.* Principles of Geographical Information Systems for land resources assessment. – Oxford University Press, 1998. – 333 p.
216. Burrough PA, Frank AU (eds) (1996) Geographic objects with indeterminate boundaries. Taylor & Francis, Basingstoke
217. Buttner G., Feranec J., Jaffrain G., et al. The CORINE Land Cover 2000 Project // EARSeL Proceedings 3(3). — 2004. — P. 331–346.

218. Canisius F. Comparison and evaluation of Medium Resolution Imaging Spectrometer leaf area index products across a range of land use / F. Canisius, R. Fernandes, J. Chen // *Remote Sens. Environ.* – 2010. – 114. - P. 950–960
219. Chabaniuk, V.S., Putrenko, V.V. Persönlicher Atlas des Ministers für Katastrophenschutz der Ukraine // *Diskussionsbeiträge zur Kartosemiotik und zur Theorie der Kartographie Intern. Korrespondenz-Seminar, Band 15.* Dresden 2012, p. 14- 19.
220. Chen J (1999) On the construction of NSDI in China. *J Remote Sens* 3(2):94–97
221. Chen J, Gong P (1998) *Practical GIS.* Science Press, Beijing
222. Chen SP (2007) A landmark masterpiece of spatial data mining: review of “Theory and Application of Spatial Data Mining”. *Science Bulletin* 52(21): 2577
223. Cheng T. Detecting emerging space-time crime patterns by prospective STSS / T. Cheng, M. Adepeju // Proc of the 12th International Conference on GeoComputation. Available: <http://www.geocomputation.org/2013/papers/77.pdf>. Accessed 2014 Jan 31
224. Chiles J. P., Delfiner P. *Geostatistics. Modeling Spatial Uncertainty.* — New York: A Wiley-Interscience Publication, 1999.
225. Cicek H. Using vegetation indices from satellite remote sensing to assess corn and soybean response to controlled tile drainage / H. Cicek, M. Sunohara, G. Wilkes, H. McNairn, F. Pick, E. Topp, D.R. Lapen // *Agric. Water Manage.* – 2010. – 98. - P. 261–270
226. Cleve C. Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography / C. Cleve, M. Kelly, F. R. Kearns, M. Morltz // *Computers Environment and Urban Systems*, 32(4). – 2008. – pp. 317–326.
227. Codd E (1995) Twelve rules for on-line analytic processing. *Computer world*, April 1995
228. conference on granular computing, pp 717–722

229. Connolly TM (2004) Database systems: a practical approach to design, implementation, and management, 4th edn. Addison-Wesley
230. Copernicus Global Land Service [Electronic resource]. – Available at: <http://land.copernicus.eu/global/products/dmp> – 08.10.2017 p.
231. Coupe T., Myck M., Najsztub M. And the Lights Went Out — Measuring the Economic Situation in Eastern Ukraine. VoxUkraine. July 18, 2016. <http://voxukraine.org/en/2016/07/18/and-the-lights-went-out-measuring-the-economic-situation-in-eastern-ukraine-en/>.
232. Craglia M, Bie K, Jackson D (2012) Digital Earth 2020: towards the vision for the next decade. *Int J Digital Earth* 5(1):4–21
233. Cressie N (1991) Statistics for spatial data. Wiley, New York
234. Cressie N. Statistics for spatial data / N. Cressie. — New York: John Wiley & Sons, 1991. — 900 p.
235. Cressie N. Statistics for spatial data. — New York: John Wiley & Sons, 1991. — 900 p.
236. Cube Application // IEEE Second International Conference on Data Stream Mining & Processing, August 21-25, 2018, Lviv, Ukraine. – pp. 79-83
237. D.J. Peuquet. It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems / D.J. Peuquet // *Annals of the Association of American Geographers*. - 84 (3). -1994. – P. 441-461.
238. David (eds) Handbook of geographic information. Springer, Berlin
239. De Cort M., Tsaturov Yu. S. Atlas on caesium contamination of Europe after the Chernobyl nuclear plant accident / European Commission. — [S. l.], 1996. — 39 p. — (Report EUR 16542 EN).
240. Defourny et al., P. Defourny, L. Schouten, S. Bartalev, S. Bontemps, P. Caccetta, A.J.W. De Wit, C. Di Bella, et al. Accuracy Assessment of a 300 M Global Land Cover Map: the GlobCover Experience (2008)
241. Di KC (2001) Spatial data mining and knowledge discovery. Wuhan University Press, Wuhan

242. Di KC, Li DR, Li DY (1997) Framework of spatial data mining and knowledge discovery. *Geomatics Inf Sci Wuhan Univ* 4:328–332
243. Eklund PW, Kirkby SD, Salim A (1998) Data mining and soil salinity analysis. *Int J Geogr Inf Sci* 12(3):247–268
244. ESRI, Help Arcmap / An overview of the Space Time Pattern Mining toolbox – Available: <http://desktop.arcgis.com/ru/arcmap/10.3/tools/space-time-pattern-mining-toolbox/an-overview-of-the-space-time-pattern-mining-toolbox.htm>
245. ESRI, Help Arcmap / Create Space Time Cube – Available: <http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm#>
246. ESRI, Help Arcmap / Create Space Time Cube – Available: <http://desktop.arcgis.com/ru/arcmap/10.3/tools/space-time-pattern-mining-toolbox/create-space-time-cube.htm>
247. Ester M et al (2000) Spatial data mining: databases primitives, algorithms and efficient DBMS support. *Data Min Knowl Disc* 4:193–216
248. Fayyad UM, Uthurusamy R (eds) (1995) Proceedings of the first international conference on knowledge discovery and data mining (KDD-95), Montreal, Canada, Aug 20–21, AAAI Press
249. Foresight 2018: systemic world conflicts and global forecast for XXI century / International Council for Science etc.; Scientific Supervisor M. Zgurovsky. – K. : NTUU «Igor Sikorsky Kyiv Polytechnic Institute», 2018. 226 p
250. Foresight of Ukrainian Economy: mid-term (2015–2020) and long-term (2020–2030) time horizons / Scientific advisor of the project acad. of NAS of Ukraine M. Zgurovsky // International Council for Science (ICSU) ; Committee for the System Analysis of the Presidium of NAS of Ukraine; National Technical University of Ukraine «Kyiv Polytechnic Institute» ; Institute for Applied System Analysis of NAS of Ukraine and MES of Ukraine ; World Data Center for Geoinformatics and Sustainable Development. — 2nd ed. — Kyiv : NTUU «KPI», Publ. house «Polytechnica», 2016. — 136 p.

251. Frasconi P, Gori M, Soda G (1999) Data categorization using decision trellises. *IEEE Trans Knowl Data Eng* 11(5):697–712
252. Fritz et al., S. Fritz, I. McCallum, C. Schill, C. Perger, R. Grillmayer, F. Achard, F. Kraxner, M. Obersteiner Geo-Wiki.Org: the use of crowdsourcing to improve global land cover *Remote Sens.*, 1 (3) (2009), pp. 345–354 <http://dx.doi.org/10.3390/rs1030345>
253. Gatalsky P. Interaction Analysis of Event Data Using Space-Time Cube / P. Gatalsky, N. Andrienko, G. Andrienko // *Proc. IEEE Eighth Int'l Conf. Information Visualization (IV '04)*. – 2004. - pp. 145-152.
254. Getis A. The Analysis of Spatial Association by Use of Distance Statistics / A. Getis, J.K. Ord // *Geographical Analysis*. – 1992. - 24(3). - P. 189-206
255. Getis, A., Ord, J.K.: The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, vol.24, no.3, (1992) 189-206
256. Giannotti F, Pedreschi D (eds) (2010) Mobility, data mining and privacy: Geographic knowledge discovery. Springer, Berlin
257. Giudici P (2003) Applied data mining: statistical methods for business and industry. Wiley, Chichester
258. Global Sustainable Development Report - 2016 edition. UN. 2016. Url: [https://sustainabledevelopment.un.org/content/documents/2328Global Sustainable development report 2016 \(final\).pdf](https://sustainabledevelopment.un.org/content/documents/2328Global%20sustainable%20development%20report%202016%20(final).pdf)
259. Gonçalves T. Visualizing Human Trajectories: Comparing Space-Time Cubes and Static Maps / T. Gonçalves , A.-P. Afonso, B. Martins // *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - 2014*. – pp. 102-107
260. Gong JY (1999) Theories and technologies on the contemporary GIS. Wuhan Technical University of Surveying and Mapping Press, Wuhan
261. Goodchild M. F. Citizens as sensors: The world of volunteered geography/ M. F. Goodchild // *Journal of Geography*, 69(4). – 2007. – P. 211–221.
262. Goodchild MF (2007) Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *Int J Spat Data Infrastruct Res* 2:24–32

263. Goodchild MF, Fu P, Rich P (2007) Sharing geographic information: an assessment of the geospatial one-stop. *Ann Assoc Am Geogr* 97(2):249–265
264. Goodin W. R., McRae G. J., Seinfeld J. H. A comparison of interpolation methods for sparse data: application to wind and concentration fields // *J. of Applied Meteorology*. — 1979. — Vol. 18. — P. 761—771.
265. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. — [S. l.]: Oxford Univ. Press, 1997.
266. Govorov M., Putrenko V., Gienko G. Mining Spatial Patterns of Distribution of Uranium in Surface and Ground Waters in Ukraine / Sami Faiz, Khaoula Mahmoudi. *Handbook of Research on Geographic Information Systems Applications and Advancements*, 2016, 567 p.
267. Govorov M., Becony   G., Gienko G., **Putrenko V.** Spatially constrained regionalization with multilayer perceptron // *Transactions in GIS*. 2019 Vol 23, Issue 4, p.1–30.
268. Grabmeier J, Rudolph A (2002) Techniques of clustering algorithms in data mining. *Data Min Knowl Disc* 6:303–360
269. Grimm J. W., Lynch J. A. Statistical analysis of error in estimating wet deposition using five surface estimation algorithms // *Atmospheric Environment*. — 1991. — Vol. 25A. — P. 317—127.
270. Grossner KE, Goodchild MF, Clarke KC (2008) Defining a digital earth system. *Trans GIS* 12(1):145–160
271. Guo D. Multivariate analysis and geovisualization with an integrated geographic knowledge discovery approach / D. Guo, M. Gahegan, A. M. MacEachren, B. Zhou // *Cartography and Geographic Information Science*, 32(2). — pp. 113–132.
272. Guo D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) / D. Guo // *International Journal of Geographical Information Science*, 22(7). — 2005. — pp. 801–823.
273. H  gerstrand T. What about people in regional science? / T. H  gerstrand // *Papers, Regional Science Association*. — 24. - 1970. — pp. 7-21.

274. Han J., Kamber M., Tung A. K. H. Spatial clustering methods in data mining: A survey / In H. J. Miller & J. Han (Eds.). – Geographic data mining and knowledge discovery. London and New York: Taylor and Francis. – 2001. – pp. 33–50
275. Han JW, Kamber M, Pei J (2012) Data mining: concepts and techniques, 3rd edn. The Morgan Kaufmann Publishers Inc, Burlington
276. Hawkins D (1980) Identifications of outliers. Chapman and Hall, London
277. Henderson J.V., Storeygard A.M., Weil D.N. Measuring Economic Growth from Outer Space. *Am. Econ. Rev.* 2012. 102(2): 994.
278. Hengl T. Finding the right pixel size // *Computers and Geosciences*. — 2006. — Vol. 32. — P. 1283—1298.
279. Hernández MA, Stolfo SJ (1998) Real-world data is dirty: data cleansing and the merge/purge problem. *Data Min Knowl Disc* 2:1–31
280. Hodge VJ, Austin J (2004) A survey of outlier detection methodologies. *Artif Intell Rev* 22(2):85–126
281. Höller J, Tsiatsis V, Mulligan C, Karnouskos S, Avesand S, Boyle D (2014) From machine-to-machine to the internet of things: introduction to a new age of intelligence. Elsevier
282. Inmon WH (2005) Building the data warehouse, 4th edn. Wiley, New York
283. J. Andrew Royle, Robert M. Dorazio. Hierarchical modeling and inference in ecology. Elsevier Ltd., 2008. - 444p.
284. Jan Feraneca, Gerard Hazeub, Susan Christensenc, Gabriel Jaffraind, Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia) // *Land Use Policy*, Volume 24, Issue 1, January 2007, Pages 234–247
285. Journel A. G., Huijbregts C. J. Mining Geostatistics. — London: Academic Press, 1978. — 600 p.
286. Journel A. G., Huijbregts Ch. J. Mining Geostatistics. — London: Academic Press, 1978. — 600 p.
287. Kanevski M. F., Arutyunyan R. V., Bolshov L. A. et al. Spatial data analysis of Chernobyl fallout data. — 1. Preliminary results / Nuclear Safety Inst. — Moscow, 1993. — 91 p. — (Preprint NSI-23-93).

288. Kanevski M., Arutyunyan R., Bolshov L. et al. Mapping of Radioactively Contaminated Territories with Geostatistics and Artificial Neural Networks // Contaminated Forests / Eds. I. Linkov and W. R. Schell. — [S. l.]: Kluwer Academic Publ., 1999. — P. 249—256.
289. Kanevski M., Demyanov V., Chernov S. et al. Geostat Office for Environmental and Pollution Spatial Data Analysis // Mathematische Geologie. — 1999. — Vol. 3, N 4. — P. 73—83.
290. Kanevski M., Maignan M. Analysis and modelling of spatial environmental data. — Lausanne: EPFL Press, 2004. — 288 p. — (With a CD and educational/research MS Windows software tools) <http://www.ppur.org/auteurs/1000772.html>).
291. Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley, New York
292. Killer J et al (1998) On combining classifier. IEEE Trans Pattern Anal Mach Intell 20(3):226–239
293. Koperski K (1999) A progressive refinement approach to spatial data mining. PhD thesis, Simon Fraser University, British Columbia
294. Koperski K. An efficient two-step method for classification of spatial data / K. Koperski, J. Han and N. Stefanovic // In 1998 international symposium on spatial data handling SDH'98, Vancouver, BC, Canada. – 1998. – pp. 45–54
295. Kostenko, T.: Types, types of military conflicts and their classification. Scientific Bulletin of the Diplomatic Academy of Ukraine. Vol. 2, (1999) 45-48 (in Ukrainian)
296. Kraak M. The Space Time Cube Revisited from a Geovisualization Perspective / M. Kraak // Proc. 21st Int'l Cartographic Conf. – 2003. - pp. 1988-1996.
297. Krige D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand // J. of the Chem., Metal. and Mining Soc. of South Africa. — 1951. — Vol. 52. — P. 119—139.

298. Kulldorff M. A spatial scan statistic / M. Kulldorff // *Communications in Statistics—Theory and Methods*, 26. – 1997. – pp. 1481–1496.
299. Lee ES (2000) Neuro-fuzzy estimation in spatial statistics. *J Math Anal Appl* 249:221–231
300. Leung Y (2009) *Knowledge discovery in spatial data*. Springer, Berlin
301. Li DR (1998) On the interpretation of “GEOMATICS”. *J Surveying Mapp* 27(2):95–98
302. Li DR, Cheng T (1994) KDG-Knowledge discovery from GIS. In: *Proceedings of the Canadian Conference on GIS*, Ottawa, Canada, June 6-10, pp 1001–1012
303. Li DR, Guan ZQ (2000) *Integration and implementation of spatial information system*. Wuhan University Press, Wuhan
304. Li DR, Wang SL, Li DY (2006) *Theories and applications of spatial data mining*. Science Press, Beijing
305. Li DR, Wang SL, Li DY (2013) *Theories and applications of spatial data mining*, 2nd edn. Science Press, Beijing
306. Li DR, Wang SL, Li DY, Wang XZ (2002) *Theories and technologies of spatial data mining and knowledge discovery*. *Geomatics Inf Sci Wuhan Univ* 27(3):221–233
307. Li DY, Du Y (2007) *Artificial intelligence with uncertainty*. Chapman and Hall/CRC, London
308. Liu B (2007) *Web data mining: exploring hyperlinks, contents, usage data*, 2nd edn. Springer, Heidelberg
309. London
310. Maceachren AM et al (1999) Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. *Int J Geogr Inf Sci* 13(4):311–334
311. Mandelbrot B. B. *The fractal theory of nature*. — New York: Freeman, 1982.
312. Mann, K.: Use ArcGIS Online to Manage Your Own Custom Map Gallery. *ArcUser*, Vol. 14, v. 4, (2011) 44-49

313. Martinez-Casasnovas J.A. Mapping multi-year cropping patterns in small irrigation districts from time-series analysis of Landsat TM images / J.A. Martinez-Casasnovas // *Eur. J. Agron.* — 2005. — 23. — P. 159—169.
314. Matheron G. Principles of Geostatistics // *Economic Geology.* — 1963. — Vol. 58. — P. 1246—1266.
315. McKinsey Global Institute (2011) Big data: the Next Frontier for Innovation, Competition, and Productivity, May 2011
316. Mennis J. Mining association rules in spatio-temporal data: An analysis of urban socioeconomic and land cover change / J. Mennis, J. W. Liu // *Transactions in GIS*, 9 (1). — 2005. — pp. 5–17.
317. Mesura V., Zavarzin A. Development and realization of agglomerative regionalization algorithms and probing of their time complexity // *Advances in electrical and computer engineering*. 2002. Vol 2(9). №1(17) 2002. - p.23-28.
318. Mesura V.I., Zavarzin A.V. ER-model of classification system of multidimensional data in vector GIS // Сборник трудов пятой научной конференции с международным участием. Болгария, Велико Търново, 16-19 октября 2002. С.53-57.
319. Mesura V.I., Zavarzin A.V. The model of informational web-site // *Development and application systems. Proceedings of the 6th International Conference.* -Suceava, Romania. 23-25 May, 2002. - P.376-379.
320. Miller H., Han J. (2009). Geographic data mining and knowledge discovery: An overview / In H. Miller & J. Han (Eds.) // *Geographic data mining and knowledge discovery*. CRC Press, Taylor and Francis Group. – 2009. – pp. 1–26.
321. Miller HJ, Han JW (eds) (2009) *Geographic data mining and knowledge discovery*, 2nd edn. The Chapman and Hall/CRC, London
322. Miller WT et al (1990) *Neural network for control*. MIT Press, Cambridge
323. Mills MP, Ottino JM (2012) The coming tech-led boom, 2012-10-12. www.wsj.com
324. Monrnonier M.S. Maximum-difference barriers: alternative numerical regionalization method. *Geogr. Anal.*, 1973, vol. 5, №3. - P.245-261.

325. Murray, Alan T. *Cluster discovery techniques for exploratory spatial data analysis [Text]* / Alan T. Murray, Vladimir Estivill-Castro // *International Journal of Geographical Information Science*. – 1998. – Vol. 12, Issue 5. – P. 431 – 443.
326. N. Pashynska, V. Snytyuk, V. Putrenko, A. Musienko A decision tree in a classification of fire hazard factors // *Eastern-European Journal of Enterprise Technologies*, 2016. - Vol 5, No 10(83) - P. 32-38
327. Naisbitt J (1991) *Megatrends 2000*. Avon Books, New York
328. National Geomatics Center of China, 2014 National Geomatics Center of China 30 M Global Land Cover Data Product (GlobeLand30) (2014)
329. Office of Science and Technology Policy (2012) Big data initiative: announces \$200 Million In New R&D Investments, March 29, 2012. www.WhiteHouse.gov/OSTP
330. Openshaw S. A mark 1 geographical analysis machine for the automated analysis of point data sets / S. Openshaw, M. Charlton, C. Wymer, A. Craft // *International Journal of Geographical Information Science*, 1(4). – 1987. – pp. 335–358.
331. Ord J.K. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application / J.K. Ord, A. Getis // *Geographical Analysis*. – 1995. – № 27(4). – pp. 13-22.
332. Pawlak Z (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, London
333. Pettorelli N. Using the satellite-derived NDVI to assess ecological responses to environmental change / N. Pettorelli, J.O. Vik, A. Mysterud, J.M. Gaillard, C.J. Tucker, N.C. Stenseth // *Trends Ecol. Evol.* - 2005. – 20. – P. 503–510.
334. Piatetsky-shapiro G (1994) An overview of knowledge discovery in databases: recent progress and challenges. In: Ziarko Wojciech P (ed) *Rough sets, fuzzy sets and knowledge discovery*. Springer, Berlin, pp 1–10
335. Pitt L, Reinke RE (1988) Criteria for polynomial time (conceptual) clustering. *Mach Learn* 2(4):371–396

336. Poh-Chin Lai, Fun-Mun So, Ka-Win g Chan. Spatial epidemiological approaches in disease mapping and analysis. — CRC Press/Tailor and Francis Group, 2009. 184p.
337. Preparata F. P., Shamos M. I. Computational Geometry. — New York: Springer-Verl., 1985. — P. 198—218.
338. Prototype Global Sustainable Development Report - Executive Summary. UN-DESA. 2013. Url: https://sustainabledevelopment.un.org/content/documents/975GSDR_Executive_Summary.pdf
339. Putrenko V .V., Pashynska N.M. Web Atlas of Sustainable Development of Ukraine // ICA Commissions Joint Workshop on Atlases, Cognition, Usability. Olomouc, April 27–30, 2018. – p. 28-29.
340. Putrenko V .V., Pashynska N.M. Wildfire prediction and monitoring in Ukraine on base of Copernicus Land service // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки. – К., С. 41-43
341. Putrenko V. Bowkun A., Spatial Modeling of the Risk of Accidents in the Power System of Ukraine // RISK Information Management, Risk Models and Applications, Berlin, Germany, November 17/18, 2014. – P. 10 – 11.
342. Putrenko V. Data mining of the land cover characteristics of Ukraine using the data of remote sensing of Earth // Slovak international scientific journal, VOL.1, №30, 2019. – p. 13-18.
343. Putrenko V. Cartographic modeling of differentiation electoral space in Ukraine // Territorial identity and geopolitics (18-23 september 2011, Kyiv-Crimea, Ukraine). – Kyiv, 2011. – P. 33 – 36
344. Putrenko V. Data Mining of Relationship in Crowdsourcing Projects and Social Activities of Citizens // 2017 IEEE First International Conference on electrical and computer engineering (UKRCON), conference proceedings, Kyiv, Ukraine, May 29 – June 2, 2017. – P. 1060 – 1066

345. Putrenko V. GIS-mapping for assessment of risks of groundwater use in Ukraine // 32nd International Geographical Congress Book of Abstracts p. 741 – 742 Режим доступу <http://igc2012.org/frontend/index.php>
346. Putrenko V. Multi-scale geospatial database for risk mapping of emergency // From cartography to geographic information science, Vilnius, Litva, May 31 – June Режим доступу: <http://www.geoportal.lt/wps/poc?uri=page:RUBRIC.2460>.
347. Putrenko V. Pashynska N. The use of remote sensing data for modeling air quality in the cities // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-5/W1, 2017, pp. 57-62
348. Putrenko V. Regional monitoring of Ukraine vegetation changes for yield forecast and climate change monitoring // Copernicus Global Land (Arlon, Belgium, 06-08 June 2016 p.). Web: <http://events.ulg.ac.be/copernicus-gl-workshop/programme/>
349. Putrenko V. Risk Management In Ukraine On The Basis Of The Functioning Of Spatial Data Infrastructure // Global Geospatial Conference 2012 - Spatially Enabling Government, Industry and Citizens, Québec City, Canada, 14-17 May 2012, Електронна публікація Режим доступу <http://www.gsdi.org/gsdiconf/gsdi13/abstracts.html#ab82>
350. Putrenko V. The Relationship Between Developing OpenStreetMap and Social Activity by the Example of Territory of Ukraine // 27th International Cartographic Conference, Maps Connecting the World, August 23-28, 2015, Rio de Janeiro / Brazil. Processing paper. Web: <http://www.icc2015.org/abstract,517.html>
351. Putrenko V., Bowkun Alexandra Spatial modeling of the risk of accident in the power system // Risk information management, risk models and applications, 2015. – CODATA-Germany LNIS. – Vol.7 - P.1 – 13.
352. Putrenko V., Govorov M., Gienko G. Unsupervised Machine Learning for Regionalization of Environmental Data: Distribution of Uranium in Groundwater in Ukraine // 27th International Cartographic Conference, Maps Connecting the

- World, August 23-28, 2015, Rio de Janeiro / Brazil. Processing paper. Web: <http://www.icc2015.org/abstract,892.html>
353. Putrenko V., Pashynska N. Analysis of regional armed conflicts using spatial clustering methods // InterCarto-InterGIS 24. Geoinformation and Sustainable Development. Processing - Bonn, July 24-28, 2018. – pp. 31-32
 354. Putrenko V., Pashynska N. Clusterization of vegetation dynamics data by method hot spot analysis // Аерокосмічні спостереження в інтересах сталого розвитку та безпеки: Матеріали доповідей. – К., 2018. – С. 42-45
 355. Putrenko V., Pashynska N. Data mining of sustainable development process with using nightlight indicators // Technology audit and production reserves — № 3/2(47), 2019. – p. 4 – 9
 356. Putrenko V., Pashynska N. Risk Modeling of Accidents in the Power System of Ukraine with Using Bayesian Network // Advances in Computer Science for Engineering and Education, Springer International Publishing AG, 2018. – vol. 754. pp.13-22
 357. Putrenko V., Pashynska N., Nazarenko S. Data Mining of Network Events With Space-Time Cube Application // First Workshop on Platial Analysis (PLATIAL'18), Heidelberg, 20–21 September 2018. – pp. 73-80
 358. Putrenko V., Rudenko L. New approaches for mapping of natural, technological and social risks and the dangers of emergency in Ukraine // 32nd International Geographical Congress Book of Abstracts p. 744 Режим доступу <http://igc2012.org/frontend/index.php>
 359. Putrenko V. Assessment of Risk and Hazard of Technological Emergencies with GIS (on the Example Vinnytsia Region) // 26th International Cartographic Conference, Dresden August 25 – 30, 2013 Proceedings. Режим доступу: <http://www.icc2013.org/?node=29>
 360. Putrenko V., Govorov M., Gienko G. Exploring Distribution of Uranium in Ukraine: Geovisualization and Spatial Statistics // 26th International Cartographic Conference, Dresden August 25 – 30, 2013 Proceedings. Режим доступу: <http://www.icc2013.org/?node=29>

361. Putrenko V., Govorov M., Gienko G. Exploring Spatial Patterns of Uranium Distribution in Ukraine // 16th AGILE Conference on Geographic Information Science: Geographic Information Science at the Heart of Europe: 14-17 May 2013, Leuven, Belgium. Режим доступа: <http://agile.gis.geo.tu-dresden.de/web/index.php/conference/proceedings>
362. Putrenko V., Modeling Spatial Relationships of Electoral Behavior in Ukraine // European review of law and international relations Chief of the Board Jerzy J. Wiatr, Warszawa, 2018. – P. 46 - 52
363. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Mateo
364. Raes F., Graziani G., Girardi F. A simple and fractal analysis of the European on-line network for airborne radioactivity monitoring // Environmental Monitoring and Assessment. — 1991. — Vol. 18. — P. 221—234.
365. Ramaswamy S, Rastogi R, Shim K (2000) Efficient algorithms for mining outliers from large data sets. In: Proceeding SIGMOD '00 proceedings of the 2000 ACM SIGMOD international conference on management of data, pp 427–438
366. Remote Sensing of Land Indicators for Sustainable Development Goal 15. NASA. 2017. Url: <https://arset.gsfc.nasa.gov/land/webinars/sdg15>
367. Reshef DN et al (2011) Detecting novel associations in large data sets. Science 334:1518
368. Revised list of global Sustainable Development Goal indicators. UN. 2016. Url: [https://unstats.un.org/sdgs/indicators/Official Revised List of global SDG indicators.pdf](https://unstats.un.org/sdgs/indicators/Official%20Revised%20List%20of%20global%20SDG%20indicators.pdf)
369. Rosenberg, J., Mateos, A.: The Cloud at Your Service. Manning Publications, (2010) 122-123
370. Saaty T.L., Sagir M. An essay on rank preservation and reversal. Mathematical and Computer Modelling. 2009. Vol.49. P.1230–1243.
371. Satellite Derived Annual PM2.5 Data Sets in Support of United Nations Sustainable Development Goals. NASA. 2017. Url: <https://arset.gsfc.nasa.gov/airquality/webinars/AQ-SDG-17>

372. Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
373. Shekhar S, Chawla S (2003) Spatial databases: a tour. Prentice Hall Inc
374. Shekhar S, Lu CT, Zhang P (2003) A unified approach to detecting spatial outliers. *GeoInformatica* 7(2):139–166
375. Shekhar S., Huang Y. Discovering spatial co-location patterns: A summary of results / In C. Jensen, M. Schneider, B. Seeger, & V. Tsotras (Eds.). // *Advances in spatial and temporal databases, proceedings, lecture notes in computer science*. Berlin: Springer-Verlag. – 2001. – pp. 236–256.
376. *Shekhar Shashi, Zhang Pusheng, Huang Yan, Vatsavai Ranga Raju* Trends in Spatial Data Mining, as a chapter to appear in *Data Mining: Next Generation Challenges and Future Directions*, H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha(eds.), AAAI/MIT Press, 2003
377. Shi WZ, Wang SL (2002) GIS attribute uncertainty and its development. *J Remote Sens* 6(5):393–400
378. Shiode S. Detection of multi-scale clusters in network space / S. Shiode, N. Shiode // *International Journal of Geographical Information Science*, 23. – 2009. – pp. 75–92.
379. Sridharan NS (ed) (1989) Proceedings of the 11th international joint conference on artificial intelligence, Detroit, MI, USA, August 20–25, Morgan Kaufmann
380. Srivastava J, Cheng PY (1999) Warehouse creation-a potential roadblock to data warehousing. *IEEE Trans Knowl Data Eng* 11(1):118–126
381. Stolte, C., Tang, D., Hanrahan, P.: Multiscale Visualization Using Data Cubes. In: *Visualization and Computer Graphics*, IEEE, (2003) 176 – 187
382. Sundberg, R., Croicu M.: UCDP GED Codebook version 5.0 Uppsala, Department of Peace and Conflict Research, Uppsala University, (2016) 39 p.
383. Sundberg, R., Melander, E.: Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, vol.50, no.4, (2013) 523-532

384. SURFER (R) Version 8.0: Surface Mapping System / Golden Software, Inc. — [S. l.], 2002.
385. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 2. Ukraine in Sustainable Development Indicators (2016– 2017). — 72 p.
386. Sustainable Development Analysis: Global and Regional Contexts / International Council for Science (ICSU) and others; Scientific Supervisor of the Project M. Zgurovsky. — K. : Igor Sikorsky KPI, 2017. — Part 1. Global Analysis of Quality and Security of Life (2016). — 208 p.
387. Tan GX (1998) The integration of spatial data structure and its indexing mechanism. *J Surveying Mapp* 27(4):293–299
388. The Applied Remote Sensing Training (ARSET) program. NASA. 2017. Url: <https://arset.gsfc.nasa.gov/>
389. The Executive Order 12906 by the President of the United States, 1994, The Harmonization of Geographic Data Access and Storage: National Spatial Data Infrastructure (NSDI) the version published by the United States Federal Register on April 13, 1994, vol 59, 71, pp 17671–17176
390. The Sustainable Development Goals Report 2016. UN. Url: <https://unstats.un.org/sdgs/report/2016/The Sustainable Development Goals Report 2016.pdf>
391. The thematic accuracy of Corine land cover 2000. Assessment using LUCAS (land use/cover area frame statistical survey) // Techn. Rept. EEA. — 2006. — N 7. — 85 p. — (www.eea.europa.eu).
392. Tung A et al. (2001) Spatial clustering in the presence of obstacles. *IEEE Transactions on Knowledge and Data Engineering*, 359–369
393. United Nations Global Pulse, 2012, Big Data for Development: Challenges & Opportunities, May 2012
394. Vapnik V. N. Statistical Learning Theory. — New York: John Wiley & Sons, Inc., 1998. — 736 p.

395. Vatsavai RR et al (2012) Spatiotemporal data mining in the era of big spatial data: algorithms and applications. In: Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data, 6–9 Nov 2012. Redondo Beach, CA, USA, pp 1–10
396. Victor Putrenko, Modeling Spatial Relationships of Electoral Behavior in Ukraine // 25th World Congress of Political Science, July 21-25, 2018, Brisbane, Australia <https://www.ipsa.org/events/congress/brisbane2018>
397. Viktor Putrenko Risk Modeling of Accidents in the Power System of Ukraine Based on SDI Data // GSDI 15 Conference Proceedings, November 2016 – p. 134- 135 – gsdiassociation.org.
398. Wackernagel H. Multivariate Geostatistics. — Berlin: Springer-Verl., 1995.
399. Wang SL (2002) Data field and cloud model based spatial data mining and knowledge discovery. Ph.D. Thesis, Wuhan University
400. Wang SL (2011) Spatial data mining under smart earth. In: Proceedings of 2011 IEEE international
401. Wang SL, Gan WY, Li DY, Li DR (2011) Data field for hierarchical clustering. *Int J Data Warehouse Min* 7(4):43–63
402. Wang SL, Shi WZ (2012) Data mining and knowledge discovery. In: Kresse Wolfgang, Danko
403. Wang SL, Yuan HN (2014) Spatial data mining: A perspective of big data. *Int J Data Warehouse Min* 10(4):50–70
404. Wang, Y.H., Chen, J.: Multi-Scale Conceptual Model for GIS Geographical Features. *Journal of China University of Mining & Technology*, Vol. 32(4), (2003) 376–382
405. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small world’ networks. *Nature* 393:400–442
406. Witten I, Frank E (2000) Data mining, practical machine learning tools and techniques with java implementation. San Francisco: Morgan Kaufman Publishers
407. Wu X, Zhu X, Wu G, Ding W (2014) Data mining with big data. *IEEE Trans Knowl Data Eng* 26(1):97–107

408. Wynne HW, Lee ML, Wang JM (2007) Temporal and spatio-temporal data mining. IGI Publishing, New York
409. Yin, Z.C.: A Multi-Scale GIS Database Model Based On Petri Net. In: ISPRS Workshop on Service and Application of Spatial Data Infrastructure, XXXVI (4/W6), Hangzhou, (2005) 271 – 275
410. Yusof N. Mining frequent spatio-temporal patterns in wind speed and direction / N. Yusof, R. Zurita-Milla, M. J. Kraak // Connecting a digital Europe through location and space (Lecture notes in geoinformation and geography). – 2014. - pp. 143-161
411. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
412. Zgurovsky M. Putrenko V., Pashynska N. .Parameterization of Sustainable Development Components Using Nightlight Indicators in Ukraine // 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), 2018. – pp. 8-12
413. Zgurovsky M.Z., Pankratova N.D. System analysis: Theory and Applications. Springer. 2007. 475 p.
414. Zgurovsky M. Z. Sustainable development global simulation: Opportunities and threats to the planet. Russian Journal of Earth Sciences, 2007, Vol. 9, ES2003, doi:10, 2205/2007ES000273 (<http://elpub.wdcb.ru/journals/rjes/v09/2007ES000273/2007ES000273.pdf>)
415. Zgurovsky M. Z. Sustainable Development Global Simulation: Analysis of Quality and Security of Human Life, 2012, DOI: 10.5772/26776 (<http://cdn.intechweb.org/pdfs/29228.pdf>)
416. Zlatanova, S.: Management of multiple representations in spatial DBMSs. In:7th AGILE Conference on Geographic Information Science, Heraklion, Greece, Parrallel Session 3, (2004) 65-67

Додаток А Індикатори моделювання сталого розвитку регіонів України
Таблиця А.1. Значення індикаторів якості життя регіонів України

Регіон	Скорочення	Назва індексу	Рік	Значення
Хмельницька	RIP	Реалізація інноваційного потенціалу	2016	0.00024242
Черкаська	RIP	Реалізація інноваційного потенціалу	2016	0.00461243
Чернівецька	RIP	Реалізація інноваційного потенціалу	2016	0.125
Чернігівська	RIP	Реалізація інноваційного потенціалу	2016	0.00451594
Вінницька	KIN	Капітальні інвестиції	2016	0.03905104
Волинська	KIN	Капітальні інвестиції	2016	0.03003043
Дніпропетровська	KIN	Капітальні інвестиції	2016	0.1560226
Донецька	KIN	Капітальні інвестиції	2016	0.05598638
Житомирська	KIN	Капітальні інвестиції	2016	0.02621701
Закарпатська	KIN	Капітальні інвестиції	2016	0.02193414
Запорізька	KIN	Капітальні інвестиції	2016	0.05192929
Івано-Франківська	KIN	Капітальні інвестиції	2016	0.03738446
Київська	KIN	Капітальні інвестиції	2016	0.15716281
Київ	KIN	Капітальні інвестиції	2016	0.5
Кіровоградська	KIN	Капітальні інвестиції	2016	0.02989449
Луганська	KIN	Капітальні інвестиції	2016	0.01939028
Львівська	KIN	Капітальні інвестиції	2016	0.0875164
Миколаївська	KIN	Капітальні інвестиції	2016	0.04576958
Донецька	ZP	Заробітна плата	2016	0.10704318
Житомирська	ZP	Заробітна плата	2016	0.07522075
Закарпатська	ZP	Заробітна плата	2016	0.08425266
Запорізька	ZP	Заробітна плата	2016	0.09220074
Івано-Франківська	ZP	Заробітна плата	2016	0.07516054
Київська	ZP	Заробітна плата	2016	0.09643068
Київ	ZP	Заробітна плата	2016	0.167
Кіровоградська	ZP	Заробітна плата	2016	0.07416703
Луганська	ZP	Заробітна плата	2016	0.09045457
Львівська	ZP	Заробітна плата	2016	0.08464404
Миколаївська	ZP	Заробітна плата	2016	0.09706292
Одеська	ZP	Заробітна плата	2016	0.09837254
Полтавська	ZP	Заробітна плата	2016	0.0853967
Рівненська	ZP	Заробітна плата	2016	0.0858483
Сумська	ZP	Заробітна плата	2016	0.07720777
Тернопільська	ZP	Заробітна плата	2016	0.07196926
Харківська	ZP	Заробітна плата	2016	0.08277745
Херсонська	ZP	Заробітна плата	2016	0.07916468
Хмельницька	ZP	Заробітна плата	2016	0.08000766
Черкаська	ZP	Заробітна плата	2016	0.07660564
Чернівецька	ZP	Заробітна плата	2016	0.07613899
Чернігівська	ZP	Заробітна плата	2016	0.07231549
Вінницька	EANA	Економічна активність населення	2016	0.1887
Волинська	EANA	Економічна активність населення	2016	0

Дніпропетровська	EANA	Економічна активність населення	2016	0.2368
Донецька	EANA	Економічна активність населення	2016	0.0222
Житомирська	EANA	Економічна активність населення	2016	0.1961
Закарпатська	EANA	Економічна активність населення	2016	0.1184
Запорізька	EANA	Економічна активність населення	2016	0.1665
Івано-Франківська	EANA	Економічна активність населення	2016	0.0851
Київська	EANA	Економічна активність населення	2016	0.1591
Київ	EANA	Економічна активність населення	2016	0.333
Кіровоградська	EANA	Економічна активність населення	2016	0.0999
Луганська	EANA	Економічна активність населення	2016	0.3145
Львівська	EANA	Економічна активність населення	2016	0.1073
Миколаївська	EANA	Економічна активність населення	2016	0.2183
Одеська	EANA	Економічна активність населення	2016	0.1147
Полтавська	EANA	Економічна активність населення	2016	0.1221
Рівненська	EANA	Економічна активність населення	2016	0.2183
Сумська	EANA	Економічна активність населення	2016	0.1813
Тернопільська	EANA	Економічна активність населення	2016	0.037
Харківська	EANA	Економічна активність населення	2016	0.2257
Херсонська	EANA	Економічна активність населення	2016	0.1887
Хмельницька	EANA	Економічна активність населення	2016	0.0666
Черкаська	EANA	Економічна активність населення	2016	0.185
Чернівецька	EANA	Економічна активність населення	2016	0.1406
Чернігівська	EANA	Економічна активність населення	2016	0.1813
Вінницька	PRP	Потреба ринку праці	2016	0.04545455
Волинська	PRP	Потреба ринку праці	2016	0.14545455
Дніпропетровська	PRP	Потреба ринку праці	2016	0.30909091
Донецька	PRP	Потреба ринку праці	2016	0.03636364
Житомирська	PRP	Потреба ринку праці	2016	0.14545455
Закарпатська	PRP	Потреба ринку праці	2016	0.03636364
Запорізька	PRP	Потреба ринку праці	2016	0.04545455
Івано-Франківська	PRP	Потреба ринку праці	2016	0.04545455
Київська	PRP	Потреба ринку праці	2016	0.20909091
Київ	PRP	Потреба ринку праці	2016	0.5
Кіровоградська	PRP	Потреба ринку праці	2016	0.13636364
Луганська	PRP	Потреба ринку праці	2016	0.03636364
Львівська	PRP	Потреба ринку праці	2016	0.25454545
Миколаївська	PRP	Потреба ринку праці	2016	0.12727273
Одеська	PRP	Потреба ринку праці	2016	0.24545455
Полтавська	PRP	Потреба ринку праці	2016	0.16363636
Рівненська	PRP	Потреба ринку праці	2016	0.08181818
Сумська	PRP	Потреба ринку праці	2016	0.09090909
Тернопільська	PRP	Потреба ринку праці	2016	0.1
Харківська	PRP	Потреба ринку праці	2016	0.25454545
Херсонська	PRP	Потреба ринку праці	2016	0.03636364
Хмельницька	PRP	Потреба ринку праці	2016	0.05454545
Черкаська	PRP	Потреба ринку праці	2016	0.04545455

Чернівецька	PRP	Потреба ринку праці	2016	0.06363636
Чернігівська	PRP	Потреба ринку праці	2016	0.06363636
Вінницька	Igr	Ринок праці	2016	0.31172359
Волинська	Igr	Ринок праці	2016	0.22299349
Дніпропетровська	Igr	Ринок праці	2016	0.63490037
Донецька	Igr	Ринок праці	2016	0.16560681
Житомирська	Igr	Ринок праці	2016	0.4167753
Закарпатська	Igr	Ринок праці	2016	0.2390163
Запорізька	Igr	Ринок праці	2016	0.30415528
Івано-Франківська	Igr	Ринок праці	2016	0.20571508
Київська	Igr	Ринок праці	2016	0.46462159
Київ	Igr	Ринок праці	2016	1
Кіровоградська	Igr	Ринок праці	2016	0.31043066
Луганська	Igr	Ринок праці	2016	0.44131821
Львівська	Igr	Ринок праці	2016	0.4464895
Миколаївська	Igr	Ринок праці	2016	0.44263564
Одеська	Igr	Ринок праці	2016	0.45852709
Полтавська	Igr	Ринок праці	2016	0.37113306
Рівненська	Igr	Ринок праці	2016	0.38596648
Сумська	Igr	Ринок праці	2016	0.34941686
Тернопільська	Igr	Ринок праці	2016	0.20896926
Харківська	Igr	Ринок праці	2016	0.5630229
Херсонська	Igr	Ринок праці	2016	0.30422832
Хмельницька	Igr	Ринок праці	2016	0.20115312
Черкаська	Igr	Ринок праці	2016	0.30706019
Чернівецька	Igr	Ринок праці	2016	0.28037536
Чернігівська	Igr	Ринок праці	2016	0.31725185
Вінницька	ID	Інноваційна діяльність	2016	0.006547
Волинська	ID	Інноваційна діяльність	2016	0.00151331
Дніпропетровська	ID	Інноваційна діяльність	2016	0.1619653
Донецька	ID	Інноваційна діяльність	2016	0.00505399
Житомирська	ID	Інноваційна діяльність	2016	0.00173
Закарпатська	ID	Інноваційна діяльність	2016	0.00189931
Запорізька	ID	Інноваційна діяльність	2016	0.02503217
Івано-Франківська	ID	Інноваційна діяльність	2016	0.00123205
Київська	ID	Інноваційна діяльність	2016	0.01272246
Київ	ID	Інноваційна діяльність	2016	0.16278011
Кіровоградська	ID	Інноваційна діяльність	2016	0.00409811
Луганська	ID	Інноваційна діяльність	2016	0.00104551
Львівська	ID	Інноваційна діяльність	2016	0.01135092
Миколаївська	ID	Інноваційна діяльність	2016	0.01661453
Одеська	ID	Інноваційна діяльність	2016	0.00916888
Полтавська	ID	Інноваційна діяльність	2016	0.00326743
Рівненська	ID	Інноваційна діяльність	2016	0.00059114
Сумська	ID	Інноваційна діяльність	2016	0.00649555
Тернопільська	ID	Інноваційна діяльність	2016	0.00107444

Харківська	ID	Інноваційна діяльність	2016	0.07117788
Херсонська	ID	Інноваційна діяльність	2016	0.00203645
Хмельницька	ID	Інноваційна діяльність	2016	0.00074758
Черкаська	ID	Інноваційна діяльність	2016	0.00385669
Чернівецька	ID	Інноваційна діяльність	2016	0.00205517
Чернігівська	ID	Інноваційна діяльність	2016	0.00177176
Вінницька	RIP	Реалізація інноваційного потенціалу	2016	0.00847096
Волинська	RIP	Реалізація інноваційного потенціалу	2016	0.0012683
Дніпропетровська	RIP	Реалізація інноваційного потенціалу	2016	0.02354707
Донецька	RIP	Реалізація інноваційного потенціалу	2016	0.062339
Житомирська	RIP	Реалізація інноваційного потенціалу	2016	0.01027754
Закарпатська	RIP	Реалізація інноваційного потенціалу	2016	0.125
Запорізька	RIP	Реалізація інноваційного потенціалу	2016	0.07911409
Івано-Франківська	RIP	Реалізація інноваційного потенціалу	2016	0.00473799
Київська	RIP	Реалізація інноваційного потенціалу	2016	0.02747048
Київ	RIP	Реалізація інноваційного потенціалу	2016	0.08895052
Кіровоградська	RIP	Реалізація інноваційного потенціалу	2016	0.00817385
Луганська	RIP	Реалізація інноваційного потенціалу	2016	0.04184782
Львівська	RIP	Реалізація інноваційного потенціалу	2016	0.04832991
Миколаївська	RIP	Реалізація інноваційного потенціалу	2016	0.00127971
Одеська	RIP	Реалізація інноваційного потенціалу	2016	0.02314625
Полтавська	RIP	Реалізація інноваційного потенціалу	2016	3.1227E-05
Рівненська	RIP	Реалізація інноваційного потенціалу	2016	0.00208258
Сумська	RIP	Реалізація інноваційного потенціалу	2016	0.14436181
Тернопільська	RIP	Реалізація інноваційного потенціалу	2016	0.00868257
Харківська	RIP	Реалізація інноваційного потенціалу	2016	0.25
Херсонська	RIP	Реалізація інноваційного потенціалу	2016	0.00116802
Чернівецька	Iksp	Індекс конкуренто-спроможності	2016	0.0725332
Чернігівська	Iksp	Індекс конкуренто-спроможності	2016	0.14508158
Вінницька	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.13136538
Волинська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.08986648
Дніпропетровська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.42954459
Донецька	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.1360476
Житомирська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.15178977
Закарпатська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.1758883
Запорізька	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.20049947
Івано-Франківська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.09206267
Київська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.2775355
Київ	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.82621144
Кіровоградська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.12264572
Луганська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.17599399
Львівська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.23698491
Миколаївська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.17735537
Одеська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.21526143
Полтавська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.16391249
Рівненська	Iikp	Індекс інноваційно-кадрового потенціалу	2016	0.13189954

Сумська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.22939977
Тернопільська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.0856161
Харківська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.4482122
Херсонська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.10862944
Хмельницька	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.09107929
Черкаська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.11944475
Чернівецька	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.18183881
Чернігівська	Iікр	Індекс інноваційно-кадрового потенціалу	2016	0.11708921
Вінницька	Iес	Індекс економічного виміру	2016	0.20075147
Волинська	Iес	Індекс економічного виміру	2016	0.10595663
Дніпропетровська	Iес	Індекс економічного виміру	2016	0.46197728
Донецька	Iес	Індекс економічного виміру	2016	0.19635305
Житомирська	Iес	Індекс економічного виміру	2016	0.14959385
Закарпатська	Iес	Індекс економічного виміру	2016	0.13829322
Запорізька	Iес	Індекс економічного виміру	2016	0.22198503
Івано-Франківська	Iес	Індекс економічного виміру	2016	0.11093577
Київська	Iес	Індекс економічного виміру	2016	0.29677655
Київ	Iес	Індекс економічного виміру	2016	0.86229095
Кіровоградська	Iес	Індекс економічного виміру	2016	0.14106028
Луганська	Iес	Індекс економічного виміру	2016	0.13076862
Львівська	Iес	Індекс економічного виміру	2016	0.24354161
Миколаївська	Iес	Індекс економічного виміру	2016	0.17756523
Одеська	Iес	Індекс економічного виміру	2016	0.25490253
Полтавська	Iес	Індекс економічного виміру	2016	0.21765305
Рівненська	Iес	Індекс економічного виміру	2016	0.12350744
Сумська	Iес	Індекс економічного виміру	2016	0.18890715
Тернопільська	Iес	Індекс економічного виміру	2016	0.10320506
Харківська	Iес	Індекс економічного виміру	2016	0.39595585
Херсонська	Iес	Індекс економічного виміру	2016	0.13087201
Хмельницька	Iес	Індекс економічного виміру	2016	0.13150421
Черкаська	Iес	Індекс економічного виміру	2016	0.16202094
Чернівецька	Iес	Індекс економічного виміру	2016	0.127186
Чернігівська	Iес	Індекс економічного виміру	2016	0.1310854
Вінницька	AIR17I	Атмосферне повітря	2016	0.57193344
Волинська	AIR17I	Атмосферне повітря	2016	0.28717522
Дніпропетровська	AIR17I	Атмосферне повітря	2016	0.109822
Донецька	AIR17I	Атмосферне повітря	2016	0.24782489
Житомирська	AIR17I	Атмосферне повітря	2016	0.7433053
Закарпатська	AIR17I	Атмосферне повітря	2016	0.38137136
Запорізька	AIR17I	Атмосферне повітря	2016	0.33690876
Івано-Франківська	AIR17I	Атмосферне повітря	2016	0.7433053
Київська	AIR17I	Атмосферне повітря	2016	0.74551276
Київ	AIR17I	Атмосферне повітря	2016	0.31985288
Кіровоградська	AIR17I	Атмосферне повітря	2016	0.68670687
Луганська	AIR17I	Атмосферне повітря	2016	0.35351587
Львівська	AIR17I	Атмосферне повітря	2016	0.56243988

Миколаївська	AIR17I	Атмосферне повітря	2016	0.27932408
Одеська	AIR17I	Атмосферне повітря	2016	0.16036067
Полтавська	AIR17I	Атмосферне повітря	2016	0.66122252
Рівненська	AIR17I	Атмосферне повітря	2016	0.48542412
Сумська	AIR17I	Атмосферне повітря	2016	0.59075599
Тернопільська	AIR17I	Атмосферне повітря	2016	0.7433053
Харківська	AIR17I	Атмосферне повітря	2016	0.72055134
Херсонська	AIR17I	Атмосферне повітря	2016	0.18991294
Хмельницька	AIR17I	Атмосферне повітря	2016	0.63655905
Черкаська	AIR17I	Атмосферне повітря	2016	0.53371099
Чернівецька	AIR17I	Атмосферне повітря	2016	0.76481126
Чернігівська	AIR17I	Атмосферне повітря	2016	0.78503652
Запорізька	Ipd	Підприємницька діяльність	2016	0.23473404
Івано-Франківська	Ipd	Підприємницька діяльність	2016	0.13640634
Київська	Ipd	Підприємницька діяльність	2016	0.27237652
Київ	Ipd	Підприємницька діяльність	2016	1
Кіровоградська	Ipd	Підприємницька діяльність	2016	0.09952751
Луганська	Ipd	Підприємницька діяльність	2016	0.07373287
Львівська	Ipd	Підприємницька діяльність	2016	0.27788478
Миколаївська	Ipd	Підприємницька діяльність	2016	0.16512009
Одеська	Ipd	Підприємницька діяльність	2016	0.32452304
Полтавська	Ipd	Підприємницька діяльність	2016	0.16092799
Рівненська	Ipd	Підприємницька діяльність	2016	0.09857436
Сумська	Ipd	Підприємницька діяльність	2016	0.10162934
Тернопільська	Ipd	Підприємницька діяльність	2016	0.09131256
Харківська	Ipd	Підприємницька діяльність	2016	0.32184634
Херсонська	Ipd	Підприємницька діяльність	2016	0.1092343
Хмельницька	Ipd	Підприємницька діяльність	2016	0.11536618
Черкаська	Ipd	Підприємницька діяльність	2016	0.12333098
Чернівецька	Ipd	Підприємницька діяльність	2016	0.06901774
Чернігівська	Ipd	Підприємницька діяльність	2016	0.09691149
Вінницька	ZP	Заробітна плата	2016	0.07756905
Волинська	ZP	Заробітна плата	2016	0.07753894
Дніпропетровська	ZP	Заробітна плата	2016	0.08900946
Одеська	KIN	Капітальні інвестиції	2016	0.0786896
Полтавська	KIN	Капітальні інвестиції	2016	0.07180502
Рівненська	KIN	Капітальні інвестиції	2016	0.02034
Сумська	KIN	Капітальні інвестиції	2016	0.02710651
Тернопільська	KIN	Капітальні інвестиції	2016	0.02299345
Харківська	KIN	Капітальні інвестиції	2016	0.07782973
Херсонська	KIN	Капітальні інвестиції	2016	0.02159687
Хмельницька	KIN	Капітальні інвестиції	2016	0.0429148
Черкаська	KIN	Капітальні інвестиції	2016	0.03056903
Чернівецька	KIN	Капітальні інвестиції	2016	0.01255368
Чернігівська	KIN	Капітальні інвестиції	2016	0.02501752
Вінницька	Iim	Інноваційно-інвестиційні можливості	2016	0.054069

Волинська	Iiim	Інноваційно-інвестиційні можливості	2016	0.03281205
Дніпропетровська	Iiim	Інноваційно-інвестиційні можливості	2016	0.34153496
Донецька	Iiim	Інноваційно-інвестиційні можливості	2016	0.12337937
Житомирська	Iiim	Інноваційно-інвестиційні можливості	2016	0.03822455
Закарпатська	Iiim	Інноваційно-інвестиційні можливості	2016	0.14883345
Запорізька	Iiim	Інноваційно-інвестиційні можливості	2016	0.15607554
Івано-Франківська	Iiim	Інноваційно-інвестиційні можливості	2016	0.0433545
Київська	Iiim	Інноваційно-інвестиційні можливості	2016	0.19735575
Київ	Iiim	Інноваційно-інвестиційні можливості	2016	0.75173063
Кіровоградська	Iiim	Інноваційно-інвестиційні можливості	2016	0.04216645
Луганська	Iiim	Інноваційно-інвестиційні можливості	2016	0.06228361
Львівська	Iiim	Інноваційно-інвестиційні можливості	2016	0.14719723
Миколаївська	Iiim	Інноваційно-інвестиційні можливості	2016	0.06366382
Одеська	Iiim	Інноваційно-інвестиційні можливості	2016	0.11100472
Полтавська	Iiim	Інноваційно-інвестиційні можливості	2016	0.07510367
Рівненська	Iiim	Інноваційно-інвестиційні можливості	2016	0.02301371
Сумська	Iiim	Інноваційно-інвестиційні можливості	2016	0.17796387
Тернопільська	Iiim	Інноваційно-інвестиційні можливості	2016	0.03275045
Харківська	Iiim	Інноваційно-інвестиційні можливості	2016	0.39900761
Херсонська	Iiim	Інноваційно-інвестиційні можливості	2016	0.02480135
Хмельницька	Iiim	Інноваційно-інвестиційні можливості	2016	0.0439048
Черкаська	Iiim	Інноваційно-інвестиційні можливості	2016	0.03903814
Чернівецька	Iiim	Інноваційно-інвестиційні можливості	2016	0.13960886
Чернігівська	Iiim	Інноваційно-інвестиційні можливості	2016	0.03130522
Вінницька	Iksp	Індекс конкуренто-спроможності	2016	0.27013757
Волинська	Iksp	Індекс конкуренто-спроможності	2016	0.12204679
Дніпропетровська	Iksp	Індекс конкуренто-спроможності	2016	0.49440997
Донецька	Iksp	Індекс конкуренто-спроможності	2016	0.2566585
Житомирська	Iksp	Індекс конкуренто-спроможності	2016	0.14739793
Закарпатська	Iksp	Індекс конкуренто-спроможності	2016	0.10069814
Запорізька	Iksp	Індекс конкуренто-спроможності	2016	0.2434706
Івано-Франківська	Iksp	Індекс конкуренто-спроможності	2016	0.12980887
Київська	Iksp	Індекс конкуренто-спроможності	2016	0.31601759
Київ	Iksp	Індекс конкуренто-спроможності	2016	0.89837046
Кіровоградська	Iksp	Індекс конкуренто-спроможності	2016	0.15947484
Луганська	Iksp	Індекс конкуренто-спроможності	2016	0.08554324
Львівська	Iksp	Індекс конкуренто-спроможності	2016	0.2500983
Миколаївська	Iksp	Індекс конкуренто-спроможності	2016	0.1777751
Одеська	Iksp	Індекс конкуренто-спроможності	2016	0.29454363
Полтавська	Iksp	Індекс конкуренто-спроможності	2016	0.2713936
Рівненська	Iksp	Індекс конкуренто-спроможності	2016	0.11511534
Сумська	Iksp	Індекс конкуренто-спроможності	2016	0.14841454
Тернопільська	Iksp	Індекс конкуренто-спроможності	2016	0.12079403
Харківська	Iksp	Індекс конкуренто-спроможності	2016	0.3436995
Херсонська	Iksp	Індекс конкуренто-спроможності	2016	0.15311458
Хмельницька	Iksp	Індекс конкуренто-спроможності	2016	0.17192912

Черкаська	Iksp	Індекс конкуренто-спроможності	2016	0.20459713
Вінницька	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.40806885
Волинська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.66241267
Дніпропетровська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.04442823
Донецька	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.06418759
Житомирська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.65251138
Закарпатська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.66446225
Запорізька	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.33694762
Івано-Франківська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.20246389
Київська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.4752337
Київ	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.63796437
Кіровоградська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.63655896
Луганська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.42384266
Львівська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.53435634
Миколаївська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.6370014
Одеська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.64038915
Полтавська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.53975387
Вінницька	LRS17I	Земельні ресурси	2016	0.33070113
Волинська	LRS17I	Земельні ресурси	2016	0.77744083
Дніпропетровська	LRS17I	Земельні ресурси	2016	0.25362453
Донецька	LRS17I	Земельні ресурси	2016	0.25218102
Житомирська	LRS17I	Земельні ресурси	2016	0.75557586
Закарпатська	LRS17I	Земельні ресурси	2016	0.91658364
Запорізька	LRS17I	Земельні ресурси	2016	0.25074296
Івано-Франківська	LRS17I	Земельні ресурси	2016	0.82583804
Київська	LRS17I	Земельні ресурси	2016	0.55085845
Київ	LRS17I	Земельні ресурси	2016	0.67676339
Кіровоградська	LRS17I	Земельні ресурси	2016	0.25218102
Луганська	LRS17I	Земельні ресурси	2016	0.42740228
Львівська	LRS17I	Земельні ресурси	2016	0.6964853
Миколаївська	LRS17I	Земельні ресурси	2016	0.25218102
Одеська	LRS17I	Земельні ресурси	2016	0.32063493
Полтавська	LRS17I	Земельні ресурси	2016	0.35478853
Рівненська	LRS17I	Земельні ресурси	2016	0.80032612
Сумська	LRS17I	Земельні ресурси	2016	0.51474085
Тернопільська	LRS17I	Земельні ресурси	2016	0.35129856
Харківська	LRS17I	Земельні ресурси	2016	0.35829402
Херсонська	LRS17I	Земельні ресурси	2016	0.3223014
Хмельницька	LRS17I	Земельні ресурси	2016	0.36535033
Черкаська	LRS17I	Земельні ресурси	2016	0.40145237
Чернівецька	LRS17I	Земельні ресурси	2016	0.6563922
Чернігівська	LRS17I	Земельні ресурси	2016	0.61574782
Вінницька	BRS17I	Заповідний фонд	2016	0.3135248
Волинська	BRS17I	Заповідний фонд	2016	0.62207808
Дніпропетровська	BRS17I	Заповідний фонд	2016	0.28195169
Донецька	BRS17I	Заповідний фонд	2016	0.40329576

Житомирська	BRS17I	Заповідний фонд	2016	0.36368897
Закарпатська	BRS17I	Заповідний фонд	2016	0.85072633
Запорізька	BRS17I	Заповідний фонд	2016	0.46527719
Івано-Франківська	BRS17I	Заповідний фонд	2016	0.7629509
Київська	BRS17I	Заповідний фонд	2016	0.74768216
Київ	BRS17I	Заповідний фонд	2016	0.89511192
Кіровоградська	BRS17I	Заповідний фонд	2016	0.27625112
Луганська	BRS17I	Заповідний фонд	2016	0.29848466
Львівська	BRS17I	Заповідний фонд	2016	0.42333117
Миколаївська	BRS17I	Заповідний фонд	2016	0.37523908
Одеська	BRS17I	Заповідний фонд	2016	0.43780228
Полтавська	BRS17I	Заповідний фонд	2016	0.31505132
Рівненська	BRS17I	Заповідний фонд	2016	0.41470959
Сумська	BRS17I	Заповідний фонд	2016	0.36314274
Тернопільська	BRS17I	Заповідний фонд	2016	0.37801074
Харківська	BRS17I	Заповідний фонд	2016	0.31149571
Херсонська	BRS17I	Заповідний фонд	2016	0.80961871
Хмельницька	BRS17I	Заповідний фонд	2016	0.89355004
Черкаська	BRS17I	Заповідний фонд	2016	0.32325982
Чернівецька	BRS17I	Заповідний фонд	2016	0.46175443
Чернігівська	BRS17I	Заповідний фонд	2016	0.34265817
Вінницька	RST17I	Радіологічний стан території та хімічна небез	2016	0.57083416
Волинська	RST17I	Радіологічний стан території та хімічна небез	2016	0.62255996
Дніпропетровська	RST17I	Радіологічний стан території та хімічна небез	2016	0.34939106
Донецька	RST17I	Радіологічний стан території та хімічна небез	2016	0.32113664
Житомирська	RST17I	Радіологічний стан території та хімічна небез	2016	0.47154096
Закарпатська	RST17I	Радіологічний стан території та хімічна небез	2016	0.52550015
Запорізька	RST17I	Радіологічний стан території та хімічна небез	2016	0.377749
Івано-Франківська	RST17I	Радіологічний стан території та хімічна небез	2016	0.49617349
Київська	RST17I	Радіологічний стан території та хімічна небез	2016	0.35212757
Київ	RST17I	Радіологічний стан території та хімічна небез	2016	0.51796825
Кіровоградська	RST17I	Радіологічний стан території та хімічна небез	2016	0.63591799
Луганська	RST17I	Радіологічний стан території та хімічна небез	2016	0.44315788
Львівська	RST17I	Радіологічний стан території та хімічна небез	2016	0.65248793
Миколаївська	RST17I	Радіологічний стан території та хімічна небез	2016	0.60136595
Одеська	RST17I	Радіологічний стан території та хімічна небез	2016	0.48542984
Полтавська	RST17I	Радіологічний стан території та хімічна небез	2016	0.54163298
Рівненська	RST17I	Радіологічний стан території та хімічна небез	2016	0.5795936
Сумська	RST17I	Радіологічний стан території та хімічна небез	2016	0.42741446
Тернопільська	RST17I	Радіологічний стан території та хімічна небез	2016	0.62826165
Харківська	RST17I	Радіологічний стан території та хімічна небез	2016	0.63342882
Херсонська	RST17I	Радіологічний стан території та хімічна небез	2016	0.63908505
Хмельницька	RST17I	Радіологічний стан території та хімічна небез	2016	0.64137695
Черкаська	RST17I	Радіологічний стан території та хімічна небез	2016	0.30883258
Чернівецька	RST17I	Радіологічний стан території та хімічна небез	2016	0.60067205
Чернігівська	RST17I	Радіологічний стан території та хімічна небез	2016	0.57614429

Вінницька	WAC17I	Відходи	2016	0.57100799
Волинська	WAC17I	Відходи	2016	0.57215374
Дніпропетровська	WAC17I	Відходи	2016	0.27458465
Донецька	WAC17I	Відходи	2016	0.42422168
Житомирська	WAC17I	Відходи	2016	0.5727896
Закарпатська	WAC17I	Відходи	2016	0.57281838
Запорізька	WAC17I	Відходи	2016	0.27830626
Івано-Франківська	WAC17I	Відходи	2016	0.56241535
Київська	WAC17I	Відходи	2016	0.56853475
Київ	WAC17I	Відходи	2016	0.54437431
Кіровоградська	WAC17I	Відходи	2016	0.54224901
Луганська	WAC17I	Відходи	2016	0.48808802
Львівська	WAC17I	Відходи	2016	0.55254953
Миколаївська	WAC17I	Відходи	2016	0.55579502
Одеська	WAC17I	Відходи	2016	0.57023972
Полтавська	WAC17I	Відходи	2016	0.56951178
Рівненська	WAC17I	Відходи	2016	0.56907581
Сумська	WAC17I	Відходи	2016	0.391311
Тернопільська	WAC17I	Відходи	2016	0.57304517
Харківська	WAC17I	Відходи	2016	0.56265906
Херсонська	WAC17I	Відходи	2016	0.57284334
Хмельницька	WAC17I	Відходи	2016	0.57232338
Черкаська	WAC17I	Відходи	2016	0.57250341
Чернівецька	WAC17I	Відходи	2016	0.57240643
Чернігівська	WAC17I	Відходи	2016	0.57242379
Вінницька	WTP17I	Водне навантаження	2016	0.62888645
Волинська	WTP17I	Водне навантаження	2016	0.6311482
Дніпропетровська	WTP17I	Водне навантаження	2016	0.28031251
Донецька	WTP17I	Водне навантаження	2016	0.22977863
Житомирська	WTP17I	Водне навантаження	2016	0.62357501
Закарпатська	WTP17I	Водне навантаження	2016	0.60934681
Запорізька	WTP17I	Водне навантаження	2016	0.37321877
Івано-Франківська	WTP17I	Водне навантаження	2016	0.62801026
Київська	WTP17I	Водне навантаження	2016	0.59311242
Київ	WTP17I	Водне навантаження	2016	0.1246126
Кіровоградська	WTP17I	Водне навантаження	2016	0.60422872
Луганська	WTP17I	Водне навантаження	2016	0.5222821
Львівська	WTP17I	Водне навантаження	2016	0.50150851
Миколаївська	WTP17I	Водне навантаження	2016	0.55560126
Одеська	WTP17I	Водне навантаження	2016	0.36813118
Полтавська	WTP17I	Водне навантаження	2016	0.62532744
Рівненська	WTP17I	Водне навантаження	2016	0.61517397
Сумська	WTP17I	Водне навантаження	2016	0.53727638
Тернопільська	WTP17I	Водне навантаження	2016	0.61972234
Харківська	WTP17I	Водне навантаження	2016	0.59222174
Херсонська	WTP17I	Водне навантаження	2016	0.33062393

Хмельницька	WTP17I	Водне навантаження	2016	0.62875526
Черкаська	WTP17I	Водне навантаження	2016	0.60584628
Чернівецька	WTP17I	Водне навантаження	2016	0.61563764
Чернігівська	WTP17I	Водне навантаження	2016	0.62039071
Вінницька	LNP17I	Утворення відходів	2016	0.60985429
Волинська	LNP17I	Утворення відходів	2016	0.61356091
Дніпропетровська	LNP17I	Утворення відходів	2016	0.23492883
Донецька	LNP17I	Утворення відходів	2016	0.28343978
Житомирська	LNP17I	Утворення відходів	2016	0.61441827
Закарпатська	LNP17I	Утворення відходів	2016	0.61065055
Запорізька	LNP17I	Утворення відходів	2016	0.56776999
Івано-Франківська	LNP17I	Утворення відходів	2016	0.59397907
Київська	LNP17I	Утворення відходів	2016	0.5608572
Київ	LNP17I	Утворення відходів	2016	0.13766402
Кіровоградська	LNP17I	Утворення відходів	2016	0.47654795
Луганська	LNP17I	Утворення відходів	2016	0.59469134
Львівська	LNP17I	Утворення відходів	2016	0.60085085
Миколаївська	LNP17I	Утворення відходів	2016	0.46257078
Одеська	LNP17I	Утворення відходів	2016	0.60262635
Полтавська	LNP17I	Утворення відходів	2016	0.49719333
Рівненська	LNP17I	Утворення відходів	2016	0.61286848
Сумська	LNP17I	Утворення відходів	2016	0.36365643
Тернопільська	LNP17I	Утворення відходів	2016	0.57511547
Харківська	LNP17I	Утворення відходів	2016	0.49905149
Херсонська	LNP17I	Утворення відходів	2016	0.5471279
Хмельницька	LNP17I	Утворення відходів	2016	0.60617623
Черкаська	LNP17I	Утворення відходів	2016	0.61059562
Чернівецька	LNP17I	Утворення відходів	2016	0.61268338
Чернігівська	LNP17I	Утворення відходів	2016	0.61397702
Вінницька	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.61424765
Волинська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.32926057
Дніпропетровська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.5456475
Донецька	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.43109645
Житомирська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.55268614
Закарпатська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.61708779
Запорізька	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.41740264
Івано-Франківська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.6161814
Київська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.24437347
Київ	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.36578769
Кіровоградська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.62831551
Луганська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.59676829
Львівська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.60723576
Миколаївська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.58913976
Одеська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.47678484
Полтавська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.64282015
Рівненська	RWE17I	Викиди забруднювальних речовин в атмосф	е 2016	0.61102482

Сумська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.58222472
Тернопільська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.43311942
Харківська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.61460953
Херсонська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.26470388
Хмельницька	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.63436065
Черкаська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.58820968
Чернівецька	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.62640486
Чернігівська	RWE17I	Викиди забруднювальних речовин в атмосфері	2016	0.42684746
Рівненська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.65132354
Сумська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.6168436
Тернопільська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.64915839
Харківська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.54801231
Херсонська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.64682789
Хмельницька	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.62045401
Черкаська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.52847779
Чернівецька	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.66641666
Чернігівська	EMP17I	Викиди забруднювальних речовин в атмосфері	2016	0.55269107
Чернівецька	EMN17C	Екологічне керування	2016	0.56798296
Чернігівська	EMN17C	Екологічне керування	2016	0.47187445
Вінницька	Ie	Індекс екологічного виміру	2016	0.50874962
Волинська	Ie	Індекс екологічного виміру	2016	0.55843539
Дніпропетровська	Ie	Індекс екологічного виміру	2016	0.34225061
Донецька	Ie	Індекс екологічного виміру	2016	0.31078098
Житомирська	Ie	Індекс екологічного виміру	2016	0.58317652
Закарпатська	Ie	Індекс екологічного виміру	2016	0.61479063
Запорізька	Ie	Індекс екологічного виміру	2016	0.39206469
Івано-Франківська	Ie	Індекс екологічного виміру	2016	0.51699772
Київська	Ie	Індекс екологічного виміру	2016	0.54239399
Київ	Ie	Індекс екологічного виміру	2016	0.45633188
Кіровоградська	Ie	Індекс екологічного виміру	2016	0.5362642
Луганська	Ie	Індекс екологічного виміру	2016	0.42327436
Львівська	Ie	Індекс екологічного виміру	2016	0.55461344
Миколаївська	Ie	Індекс екологічного виміру	2016	0.5059098
Одеська	Ie	Індекс екологічного виміру	2016	0.48038389
Полтавська	Ie	Індекс екологічного виміру	2016	0.53974456
Рівненська	Ie	Індекс екологічного виміру	2016	0.57838454
Сумська	Ie	Індекс екологічного виміру	2016	0.51771381
Тернопільська	Ie	Індекс екологічного виміру	2016	0.55469488
Харківська	Ie	Індекс екологічного виміру	2016	0.53092832
Херсонська	Ie	Індекс екологічного виміру	2016	0.48872061
Хмельницька	Ie	Індекс екологічного виміру	2016	0.5909222
Черкаська	Ie	Індекс екологічного виміру	2016	0.50981059
Чернівецька	Ie	Індекс екологічного виміру	2016	0.60358982
Чернігівська	Ie	Індекс екологічного виміру	2016	0.54865428
Вінницька	CCM17I	Змінювання клімату	2016	0.49004828
Волинська	CCM17I	Змінювання клімату	2016	0.64896505

Дніпропетровська	CCM17I	Змінювання клімату	2016	0.32480088
Донецька	CCM17I	Змінювання клімату	2016	0.07437185
Житомирська	CCM17I	Змінювання клімату	2016	0.64776725
Закарпатська	CCM17I	Змінювання клімату	2016	0.67232663
Запорізька	CCM17I	Змінювання клімату	2016	0.32462405
Івано-Франківська	CCM17I	Змінювання клімату	2016	0.13849786
Київська	CCM17I	Змінювання клімату	2016	0.57646213
Київ	CCM17I	Змінювання клімату	2016	0.65739002
Кіровоградська	CCM17I	Змінювання клімату	2016	0.62469207
Луганська	CCM17I	Змінювання клімату	2016	0.05053564
Львівська	CCM17I	Змінювання клімату	2016	0.60168733
Миколаївська	CCM17I	Змінювання клімату	2016	0.58607119
Одеська	CCM17I	Змінювання клімату	2016	0.63739832
Полтавська	CCM17I	Змінювання клімату	2016	0.60387666
Рівненська	CCM17I	Змінювання клімату	2016	0.59432911
Сумська	CCM17I	Змінювання клімату	2016	0.58932132
Тернопільська	CCM17I	Змінювання клімату	2016	0.6493648
Харківська	CCM17I	Змінювання клімату	2016	0.55487486
Херсонська	CCM17I	Змінювання клімату	2016	0.66089386
Хмельницька	CCM17I	Змінювання клімату	2016	0.54759045
Черкаська	CCM17I	Змінювання клімату	2016	0.54854514
Чернівецька	CCM17I	Змінювання клімату	2016	0.66546352
Чернігівська	CCM17I	Змінювання клімату	2016	0.57647081
Вінницька	WSU17I	Керування відходами	2016	0.41284016
Волинська	WSU17I	Керування відходами	2016	0.41172642
Дніпропетровська	WSU17I	Керування відходами	2016	0.88851317
Донецька	WSU17I	Керування відходами	2016	0.72495848
Житомирська	WSU17I	Керування відходами	2016	0.41349024
Закарпатська	WSU17I	Керування відходами	2016	0.41104173
Запорізька	WSU17I	Керування відходами	2016	0.48322862
Івано-Франківська	WSU17I	Керування відходами	2016	0.43943715
Київська	WSU17I	Керування відходами	2016	0.65267756
Київ	WSU17I	Керування відходами	2016	0.41112581
Кіровоградська	WSU17I	Керування відходами	2016	0.420051
Луганська	WSU17I	Керування відходами	2016	0.41495999
Львівська	WSU17I	Керування відходами	2016	0.41350587
Миколаївська	WSU17I	Керування відходами	2016	0.48446073
Одеська	WSU17I	Керування відходами	2016	0.41344473
Полтавська	WSU17I	Керування відходами	2016	0.48340533
Рівненська	WSU17I	Керування відходами	2016	0.41126341
Сумська	WSU17I	Керування відходами	2016	0.5976828
Тернопільська	WSU17I	Керування відходами	2016	0.46142069
Харківська	WSU17I	Керування відходами	2016	0.41772754
Херсонська	WSU17I	Керування відходами	2016	0.42805118
Хмельницька	WSU17I	Керування відходами	2016	0.41546736
Черкаська	WSU17I	Керування відходами	2016	0.42276532

Чернівецька	WSU17I	Керування відходами	2016	0.41208049
Чернігівська	WSU17I	Керування відходами	2016	0.41230506
Вінницька	EST17C	Стан довкілля	2016	0.4716003
Волинська	EST17C	Стан довкілля	2016	0.57628157
Дніпропетровська	EST17C	Стан довкілля	2016	0.25387479
Донецька	EST17C	Стан довкілля	2016	0.329732
Житомирська	EST17C	Стан довкілля	2016	0.58138014
Закарпатська	EST17C	Стан довкілля	2016	0.64939997
Запорізька	EST17C	Стан довкілля	2016	0.34179683
Івано-Франківська	EST17C	Стан довкілля	2016	0.67813662
Київська	EST17C	Стан довкілля	2016	0.59294314
Київ	EST17C	Стан довкілля	2016	0.59081415
Кіровоградська	EST17C	Стан довкілля	2016	0.4786612
Луганська	EST17C	Стан довкілля	2016	0.40212974
Львівська	EST17C	Стан довкілля	2016	0.57745876
Миколаївська	EST17C	Стан довкілля	2016	0.41278103
Одеська	EST17C	Стан довкілля	2016	0.39489349
Полтавська	EST17C	Стан довкілля	2016	0.48844142
Рівненська	EST17C	Стан довкілля	2016	0.56982585
Сумська	EST17C	Стан довкілля	2016	0.45747301
Тернопільська	EST17C	Стан довкілля	2016	0.53478428
Харківська	EST17C	Стан довкілля	2016	0.51728579
Херсонська	EST17C	Стан довкілля	2016	0.50675229
Хмельницька	EST17C	Стан довкілля	2016	0.62183195
Черкаська	EST17C	Стан довкілля	2016	0.42795183
Чернівецька	EST17C	Стан довкілля	2016	0.61120727
Чернігівська	EST17C	Стан довкілля	2016	0.57840212
Вінницька	EPR17C	Екологічне навантаження і небезпеки	2016	0.54893653
Волинська	EPR17C	Екологічне навантаження і небезпеки	2016	0.63570726
Дніпропетровська	EPR17C	Екологічне навантаження і небезпеки	2016	0.18655652
Донецька	EPR17C	Екологічне навантаження і небезпеки	2016	0.19246867
Житомирська	EPR17C	Екологічне навантаження і небезпеки	2016	0.63016822
Закарпатська	EPR17C	Екологічне навантаження і небезпеки	2016	0.6281532
Запорізька	EPR17C	Екологічне навантаження і небезпеки	2016	0.42597879
Івано-Франківська	EPR17C	Екологічне навантаження і небезпеки	2016	0.47481774
Київська	EPR17C	Екологічне навантаження і небезпеки	2016	0.54306778
Київ	EPR17C	Екологічне навантаження і небезпеки	2016	0.30008033
Кіровоградська	EPR17C	Екологічне навантаження і небезпеки	2016	0.57244521
Луганська	EPR17C	Екологічне навантаження і небезпеки	2016	0.51360537
Львівська	EPR17C	Екологічне навантаження і небезпеки	2016	0.5455719
Миколаївська	EPR17C	Екологічне навантаження і небезпеки	2016	0.55172448
Одеська	EPR17C	Екологічне навантаження і небезпеки	2016	0.53704889
Полтавська	EPR17C	Екологічне навантаження і небезпеки	2016	0.55409155
Рівненська	EPR17C	Екологічне навантаження і небезпеки	2016	0.62645533
Сумська	EPR17C	Екологічне навантаження і небезпеки	2016	0.50592547
Тернопільська	EPR17C	Екологічне навантаження і небезпеки	2016	0.6146654

Харківська	EPR17C	Екологічне навантаження і небезпеки	2016	0.54642852
Херсонська	EPR17C	Екологічне навантаження і небезпеки	2016	0.50819324
Хмельницька	EPR17C	Екологічне навантаження і небезпеки	2016	0.61846183
Черкаська	EPR17C	Екологічне навантаження і небезпеки	2016	0.5816399
Чернівецька	EPR17C	Екологічне навантаження і небезпеки	2016	0.63157923
Чернігівська	EPR17C	Екологічне навантаження і небезпеки	2016	0.59568627
Вінницька	EMN17C	Екологічне керування	2016	0.50571203
Волинська	EMN17C	Екологічне керування	2016	0.46331735
Дніпропетровська	EMN17C	Екологічне керування	2016	0.58632052
Донецька	EMN17C	Екологічне керування	2016	0.41014226
Житомирська	EMN17C	Екологічне керування	2016	0.53798121
Закарпатська	EMN17C	Екологічне керування	2016	0.56681872
Запорізька	EMN17C	Екологічне керування	2016	0.40841844
Івано-Франківська	EMN17C	Екологічне керування	2016	0.3980388
Київська	EMN17C	Екологічне керування	2016	0.49117105
Київ	EMN17C	Екологічне керування	2016	0.47810117
Кіровоградська	EMN17C	Екологічне керування	2016	0.55768619
Луганська	EMN17C	Екологічне керування	2016	0.35408797
Львівська	EMN17C	Екологічне керування	2016	0.54080965
Миколаївська	EMN17C	Екологічне керування	2016	0.55322389
Одеська	EMN17C	Екологічне керування	2016	0.5092093
Полтавська	EMN17C	Екологічне керування	2016	0.57670071
Рівненська	EMN17C	Екологічне керування	2016	0.53887245
Сумська	EMN17C	Екологічне керування	2016	0.58974295
Тернопільська	EMN17C	Екологічне керування	2016	0.51463497
Харківська	EMN17C	Екологічне керування	2016	0.52907065
Херсонська	EMN17C	Екологічне керування	2016	0.4512163
Хмельницька	EMN17C	Екологічне керування	2016	0.53247282
Черкаська	EMN17C	Екологічне керування	2016	0.51984005
Вінницька	HLT	Розвиток здоров'я	2016	0.60328738
Волинська	HLT	Розвиток здоров'я	2016	0.53547676
Дніпропетровська	HLT	Розвиток здоров'я	2016	0.33131499
Донецька	HLT	Розвиток здоров'я	2016	0.49890511
Житомирська	HLT	Розвиток здоров'я	2016	0.49955904
Закарпатська	HLT	Розвиток здоров'я	2016	0.56277878
Запорізька	HLT	Розвиток здоров'я	2016	0.46484809
Івано-Франківська	HLT	Розвиток здоров'я	2016	0.60918005
Київська	HLT	Розвиток здоров'я	2016	0.39272508
Київ	HLT	Розвиток здоров'я	2016	0.57502995
Кіровоградська	HLT	Розвиток здоров'я	2016	0.47175622
Луганська	HLT	Розвиток здоров'я	2016	0.56449283
Львівська	HLT	Розвиток здоров'я	2016	0.5464449
Миколаївська	HLT	Розвиток здоров'я	2016	0.37795246
Одеська	HLT	Розвиток здоров'я	2016	0.20945303
Полтавська	HLT	Розвиток здоров'я	2016	0.54371577
Рівненська	HLT	Розвиток здоров'я	2016	0.39846748

Сумська	HLT	Розвиток здоров'я	2016	0.57597656
Тернопільська	HLT	Розвиток здоров'я	2016	0.64536713
Харківська	HLT	Розвиток здоров'я	2016	0.62908785
Херсонська	HLT	Розвиток здоров'я	2016	0.41879552
Хмельницька	HLT	Розвиток здоров'я	2016	0.5759295
Черкаська	HLT	Розвиток здоров'я	2016	0.53411669
Чернівецька	HLT	Розвиток здоров'я	2016	0.69687248
Чернігівська	HLT	Розвиток здоров'я	2016	0.46280699
Вінницька	Ihd	Індекс людського розвитку	2016	0.67683622
Волинська	Ihd	Індекс людського розвитку	2016	0.66411303
Дніпропетровська	Ihd	Індекс людського розвитку	2016	0.6092845
Донецька	Ihd	Індекс людського розвитку	2016	0.64917744
Житомирська	Ihd	Індекс людського розвитку	2016	0.63772771
Закарпатська	Ihd	Індекс людського розвитку	2016	0.65595363
Запорізька	Ihd	Індекс людського розвитку	2016	0.65088443
Івано-Франківська	Ihd	Індекс людського розвитку	2016	0.68557401
Київська	Ihd	Індекс людського розвитку	2016	0.64281752
Київ	Ihd	Індекс людського розвитку	2016	0.67201489
Кіровоградська	Ihd	Індекс людського розвитку	2016	0.62948687
Луганська	Ihd	Індекс людського розвитку	2016	0.66885376
Львівська	Ihd	Індекс людського розвитку	2016	0.67891347
Миколаївська	Ihd	Індекс людського розвитку	2016	0.61127574
Одеська	Ihd	Індекс людського розвитку	2016	0.57310591
Полтавська	Ihd	Індекс людського розвитку	2016	0.66806473
Рівненська	Ihd	Індекс людського розвитку	2016	0.60210024
Сумська	Ihd	Індекс людського розвитку	2016	0.67377297
Тернопільська	Ihd	Індекс людського розвитку	2016	0.68795014
Харківська	Ihd	Індекс людського розвитку	2016	0.70396635
Херсонська	Ihd	Індекс людського розвитку	2016	0.61961866
Хмельницька	Ihd	Індекс людського розвитку	2016	0.66516885
Черкаська	Ihd	Індекс людського розвитку	2016	0.65728501
Чернівецька	Ihd	Індекс людського розвитку	2016	0.69962175
Чернігівська	Ihd	Індекс людського розвитку	2016	0.6213821
Вінницька	PCR	Ефективність державної влади	2016	0.51558504
Волинська	PCR	Ефективність державної влади	2016	0.51988066
Дніпропетровська	PCR	Ефективність державної влади	2016	0.31556321
Донецька	PCR	Ефективність державної влади	2016	0.39810065
Житомирська	PCR	Ефективність державної влади	2016	0.47639034
Закарпатська	PCR	Ефективність державної влади	2016	0.51437956
Запорізька	PCR	Ефективність державної влади	2016	0.51077672
Івано-Франківська	PCR	Ефективність державної влади	2016	0.54246866
Київська	PCR	Ефективність державної влади	2016	0.48386009
Київ	PCR	Ефективність державної влади	2016	0.46248717
Кіровоградська	PCR	Ефективність державної влади	2016	0.43438262
Луганська	PCR	Ефективність державної влади	2016	0.39496389
Львівська	PCR	Ефективність державної влади	2016	0.46018164

Миколаївська	PCP	Ефективність державної влади	2016	0.49146914
Одеська	PCP	Ефективність державної влади	2016	0.53348209
Полтавська	PCP	Ефективність державної влади	2016	0.52019976
Рівненська	PCP	Ефективність державної влади	2016	0.5507852
Сумська	PCP	Ефективність державної влади	2016	0.40190251
Тернопільська	PCP	Ефективність державної влади	2016	0.52899547
Харківська	PCP	Ефективність державної влади	2016	0.42239921
Херсонська	PCP	Ефективність державної влади	2016	0.69561912
Хмельницька	PCP	Ефективність державної влади	2016	0.62867812
Черкаська	PCP	Ефективність державної влади	2016	0.57720509
Чернівецька	PCP	Ефективність державної влади	2016	0.49540399
Чернігівська	PCP	Ефективність державної влади	2016	0.48590324
Вінницька	DCS	Розвиток інститутів громадянського суспільства	2016	0.45682768
Волинська	DCS	Розвиток інститутів громадянського суспільства	2016	0.63346981
Дніпропетровська	DCS	Розвиток інститутів громадянського суспільства	2016	0.34400033
Донецька	DCS	Розвиток інститутів громадянського суспільства	2016	0.4537446
Житомирська	DCS	Розвиток інститутів громадянського суспільства	2016	0.44300097
Закарпатська	DCS	Розвиток інститутів громадянського суспільства	2016	0.58531321
Запорізька	DCS	Розвиток інститутів громадянського суспільства	2016	0.38791846
Івано-Франківська	DCS	Розвиток інститутів громадянського суспільства	2016	0.57324028
Вінницька	DIS	Розвиток інформаційного суспільства	2016	0.57934599
Волинська	DIS	Розвиток інформаційного суспільства	2016	0.55811448
Дніпропетровська	DIS	Розвиток інформаційного суспільства	2016	0.47741406
Донецька	DIS	Розвиток інформаційного суспільства	2016	0.31496366
Житомирська	DIS	Розвиток інформаційного суспільства	2016	0.55086178
Закарпатська	DIS	Розвиток інформаційного суспільства	2016	0.65197539
Запорізька	DIS	Розвиток інформаційного суспільства	2016	0.52936648
Івано-Франківська	DIS	Розвиток інформаційного суспільства	2016	0.56350072
Київська	DIS	Розвиток інформаційного суспільства	2016	0.51693245
Київ	DIS	Розвиток інформаційного суспільства	2016	0.64338179
Кіровоградська	DIS	Розвиток інформаційного суспільства	2016	0.34467887
Луганська	DIS	Розвиток інформаційного суспільства	2016	0.35789175
Львівська	DIS	Розвиток інформаційного суспільства	2016	0.49266269
Миколаївська	DIS	Розвиток інформаційного суспільства	2016	0.46097101
Одеська	DIS	Розвиток інформаційного суспільства	2016	0.51269169
Полтавська	DIS	Розвиток інформаційного суспільства	2016	0.41770274
Рівненська	DIS	Розвиток інформаційного суспільства	2016	0.51095317
Сумська	DIS	Розвиток інформаційного суспільства	2016	0.33791285
Тернопільська	DIS	Розвиток інформаційного суспільства	2016	0.52682263
Харківська	DIS	Розвиток інформаційного суспільства	2016	0.47990499
Херсонська	DIS	Розвиток інформаційного суспільства	2016	0.45313746
Хмельницька	DIS	Розвиток інформаційного суспільства	2016	0.53845998
Черкаська	DIS	Розвиток інформаційного суспільства	2016	0.51442953
Чернівецька	DIS	Розвиток інформаційного суспільства	2016	0.3170111
Чернігівська	DIS	Розвиток інформаційного суспільства	2016	0.45270491
Київська	DCS	Розвиток інститутів громадянського суспільства	2016	0.4444003

Київ	DCS	Розвиток інститутів громадянського суспільства	2016	0.74588094
Кіровоградська	DCS	Розвиток інститутів громадянського суспільства	2016	0.44545074
Луганська	DCS	Розвиток інститутів громадянського суспільства	2016	0.5668676
Львівська	DCS	Розвиток інститутів громадянського суспільства	2016	0.59050396
Миколаївська	DCS	Розвиток інститутів громадянського суспільства	2016	0.38363056
Одеська	DCS	Розвиток інститутів громадянського суспільства	2016	0.39222914
Полтавська	DCS	Розвиток інститутів громадянського суспільства	2016	0.43043162
Рівненська	DCS	Розвиток інститутів громадянського суспільства	2016	0.56958878
Сумська	DCS	Розвиток інститутів громадянського суспільства	2016	0.45813752
Тернопільська	DCS	Розвиток інститутів громадянського суспільства	2016	0.56860684
Харківська	DCS	Розвиток інститутів громадянського суспільства	2016	0.44874146
Херсонська	DCS	Розвиток інститутів громадянського суспільства	2016	0.28037882
Хмельницька	DCS	Розвиток інститутів громадянського суспільства	2016	0.57428655
Черкаська	DCS	Розвиток інститутів громадянського суспільства	2016	0.45652485
Чернівецька	DCS	Розвиток інститутів громадянського суспільства	2016	0.57021286
Чернігівська	DCS	Розвиток інститутів громадянського суспільства	2016	0.42845391
Івано-Франківська	RS	Ризики та безпека життя	2016	0.4372697
Київська	RS	Ризики та безпека життя	2016	0.39867336
Київ	RS	Ризики та безпека життя	2016	0.10949577
Кіровоградська	RS	Ризики та безпека життя	2016	0.517612
Вінницька	REL	Вплив релігійних інституцій	2016	0.67684293
Волинська	REL	Вплив релігійних інституцій	2016	0.58533663
Дніпропетровська	REL	Вплив релігійних інституцій	2016	0.46920187
Донецька	REL	Вплив релігійних інституцій	2016	0.44693523
Житомирська	REL	Вплив релігійних інституцій	2016	0.61195381
Закарпатська	REL	Вплив релігійних інституцій	2016	0.76309495
Запорізька	REL	Вплив релігійних інституцій	2016	0.48813741
Івано-Франківська	REL	Вплив релігійних інституцій	2016	0.57789871
Київська	REL	Вплив релігійних інституцій	2016	0.51220728
Київ	REL	Вплив релігійних інституцій	2016	0.45431019
Кіровоградська	REL	Вплив релігійних інституцій	2016	0.41975454
Луганська	REL	Вплив релігійних інституцій	2016	0.479138
Львівська	REL	Вплив релігійних інституцій	2016	0.81166827
Миколаївська	REL	Вплив релігійних інституцій	2016	0.32932708
Одеська	REL	Вплив релігійних інституцій	2016	0.52249165
Полтавська	REL	Вплив релігійних інституцій	2016	0.24468473
Рівненська	REL	Вплив релігійних інституцій	2016	0.65275122
Сумська	REL	Вплив релігійних інституцій	2016	0.13517366
Тернопільська	REL	Вплив релігійних інституцій	2016	0.58733652
Харківська	REL	Вплив релігійних інституцій	2016	0.28853209
Херсонська	REL	Вплив релігійних інституцій	2016	0.50263375
Хмельницька	REL	Вплив релігійних інституцій	2016	0.65928202
Черкаська	REL	Вплив релігійних інституцій	2016	0.51091554
Чернівецька	REL	Вплив релігійних інституцій	2016	0.58228933
Чернігівська	REL	Вплив релігійних інституцій	2016	0.23380347
Вінницька	IID	Інституціональний розвиток	2016	0.53614977

Волинська	IID	Інституціональний розвиток	2016	0.56629369
Дніпропетровська	IID	Інституціональний розвиток	2016	0.38659161
Донецька	IID	Інституціональний розвиток	2016	0.391954
Житомирська	IID	Інституціональний розвиток	2016	0.50394078
Закарпатська	IID	Інституціональний розвиток	2016	0.59826326
Запорізька	IID	Інституціональний розвиток	2016	0.48337515
Івано-Франківська	IID	Інституціональний розвиток	2016	0.56001419
Київська	IID	Інституціональний розвиток	2016	0.48675157
Київ	IID	Інституціональний розвиток	2016	0.5867863
Кіровоградська	IID	Інституціональний розвиток	2016	0.40877572
Луганська	IID	Інституціональний розвиток	2016	0.43523558
Львівська	IID	Інституціональний розвиток	2016	0.5376552
Миколаївська	IID	Інституціональний розвиток	2016	0.43914585
Одеська	IID	Інституціональний розвиток	2016	0.49083269
Полтавська	IID	Інституціональний розвиток	2016	0.43945711
Рівненська	IID	Інституціональний розвиток	2016	0.55373309
Сумська	IID	Інституціональний розвиток	2016	0.37009148
Тернопільська	IID	Інституціональний розвиток	2016	0.54408057
Харківська	IID	Інституціональний розвиток	2016	0.4328498
Херсонська	IID	Інституціональний розвиток	2016	0.49976601
Хмельницька	IID	Інституціональний розвиток	2016	0.59107517
Черкаська	IID	Інституціональний розвиток	2016	0.52157341
Чернівецька	IID	Інституціональний розвиток	2016	0.46927687
Чернігівська	IID	Інституціональний розвиток	2016	0.43637143
Вінницька	INF	Стан соціальної інфраструктури	2016	0.53807721
Волинська	INF	Стан соціальної інфраструктури	2016	0.37711553
Дніпропетровська	INF	Стан соціальної інфраструктури	2016	0.65413084
Донецька	INF	Стан соціальної інфраструктури	2016	0.49770158
Житомирська	INF	Стан соціальної інфраструктури	2016	0.46119449
Закарпатська	INF	Стан соціальної інфраструктури	2016	0.3683925
Запорізька	INF	Стан соціальної інфраструктури	2016	0.57350533
Івано-Франківська	INF	Стан соціальної інфраструктури	2016	0.43111842
Київська	INF	Стан соціальної інфраструктури	2016	0.50065903
Київ	INF	Стан соціальної інфраструктури	2016	0.72870861
Кіровоградська	INF	Стан соціальної інфраструктури	2016	0.41422033
Луганська	INF	Стан соціальної інфраструктури	2016	0.36951568
Львівська	INF	Стан соціальної інфраструктури	2016	0.52560928
Миколаївська	INF	Стан соціальної інфраструктури	2016	0.52837717
Одеська	INF	Стан соціальної інфраструктури	2016	0.59756316
Полтавська	INF	Стан соціальної інфраструктури	2016	0.52453722
Рівненська	INF	Стан соціальної інфраструктури	2016	0.36568356
Сумська	INF	Стан соціальної інфраструктури	2016	0.47074819
Тернопільська	INF	Стан соціальної інфраструктури	2016	0.41227575
Харківська	INF	Стан соціальної інфраструктури	2016	0.65067278
Херсонська	INF	Стан соціальної інфраструктури	2016	0.4927134
Хмельницька	INF	Стан соціальної інфраструктури	2016	0.46153675

Черкаська	INF	Стан соціальної інфраструктури	2016	0.48245711
Чернівецька	INF	Стан соціальної інфраструктури	2016	0.42354084
Чернігівська	INF	Стан соціальної інфраструктури	2016	0.45040675
Вінницька	RS	Ризики та безпека життя	2016	0.56115569
Волинська	RS	Ризики та безпека життя	2016	0.68065385
Дніпропетровська	RS	Ризики та безпека життя	2016	0.4219227
Донецька	RS	Ризики та безпека життя	2016	0.40612688
Житомирська	RS	Ризики та безпека життя	2016	0.52467265
Закарпатська	RS	Ризики та безпека життя	2016	0.67519424
Запорізька	RS	Ризики та безпека життя	2016	0.44279228
Вінницька	EW	Економічне благополуччя	2016	0.56544672
Волинська	EW	Економічне благополуччя	2016	0.37069284
Дніпропетровська	EW	Економічне благополуччя	2016	0.4619474
Донецька	EW	Економічне благополуччя	2016	0.35493944
Житомирська	EW	Економічне благополуччя	2016	0.55104098
Закарпатська	EW	Економічне благополуччя	2016	0.58227284
Запорізька	EW	Економічне благополуччя	2016	0.54734613
Івано-Франківська	EW	Економічне благополуччя	2016	0.51229811
Київська	EW	Економічне благополуччя	2016	0.54119542
Київ	EW	Економічне благополуччя	2016	0.66861501
Кіровоградська	EW	Економічне благополуччя	2016	0.49736822
Луганська	EW	Економічне благополуччя	2016	0.51172713
Львівська	EW	Економічне благополуччя	2016	0.50308384
Миколаївська	EW	Економічне благополуччя	2016	0.63111189
Одеська	EW	Економічне благополуччя	2016	0.53511366
Полтавська	EW	Економічне благополуччя	2016	0.4780487
Рівненська	EW	Економічне благополуччя	2016	0.45949735
Сумська	EW	Економічне благополуччя	2016	0.47883862
Тернопільська	EW	Економічне благополуччя	2016	0.42861848
Харківська	EW	Економічне благополуччя	2016	0.50586198
Херсонська	EW	Економічне благополуччя	2016	0.50736122
Хмельницька	EW	Економічне благополуччя	2016	0.40223319
Черкаська	EW	Економічне благополуччя	2016	0.42770556
Чернівецька	EW	Економічне благополуччя	2016	0.56994497
Чернігівська	EW	Економічне благополуччя	2016	0.54053883
Вінницька	IqI	Індекс якості життя	2016	0.55436553
Волинська	IqI	Індекс якості життя	2016	0.48142714
Дніпропетровська	IqI	Індекс якості життя	2016	0.51520296
Донецька	IqI	Індекс якості життя	2016	0.42282179
Житомирська	IqI	Індекс якості життя	2016	0.51036579
Закарпатська	IqI	Індекс якості життя	2016	0.53993721
Запорізька	IqI	Індекс якості життя	2016	0.519908
Івано-Франківська	IqI	Індекс якості життя	2016	0.45762527
Київська	IqI	Індекс якості життя	2016	0.47712496
Київ	IqI	Індекс якості життя	2016	0.49395603
Кіровоградська	IqI	Індекс якості життя	2016	0.47535178

Вінницька	Sec	Економічна безпека	2016	0.70030672
Волинська	Sec	Економічна безпека	2016	0.77504252
Дніпропетровська	Sec	Економічна безпека	2016	0.7578981
Донецька	Sec	Економічна безпека	2016	0.91018515
Житомирська	Sec	Економічна безпека	2016	0.77149323
Закарпатська	Sec	Економічна безпека	2016	0.6612457
Запорізька	Sec	Економічна безпека	2016	0.80145429
Івано-Франківська	Sec	Економічна безпека	2016	0.63562384
Київська	Sec	Економічна безпека	2016	0.69384819
Київ	Sec	Економічна безпека	2016	1.02886685
Кіровоградська	Sec	Економічна безпека	2016	0.8476374
Луганська	Sec	Економічна безпека	2016	0.97445896
Львівська	Sec	Економічна безпека	2016	0.64475469
Миколаївська	Sec	Економічна безпека	2016	0.70066167
Одеська	Sec	Економічна безпека	2016	0.68119749
Полтавська	Sec	Економічна безпека	2016	0.89100339
Рівненська	Sec	Економічна безпека	2016	0.72235484
Сумська	Sec	Економічна безпека	2016	0.6942955
Тернопільська	Sec	Економічна безпека	2016	0.76711865
Харківська	Sec	Економічна безпека	2016	0.66739731
Херсонська	Sec	Економічна безпека	2016	0.77142199
Хмельницька	Sec	Економічна безпека	2016	0.68208113
Черкаська	Sec	Економічна безпека	2016	0.71687047
Чернівецька	Sec	Економічна безпека	2016	0.60724809
Чернігівська	Sec	Економічна безпека	2016	0.7795883
Луганська	RS	Ризики та безпека життя	2016	0.69537874
Львівська	RS	Ризики та безпека життя	2016	0.42028131
Миколаївська	RS	Ризики та безпека життя	2016	0.62726423
Одеська	RS	Ризики та безпека життя	2016	0.30928578
Полтавська	RS	Ризики та безпека життя	2016	0.51782652
Рівненська	RS	Ризики та безпека життя	2016	0.61162139
Сумська	RS	Ризики та безпека життя	2016	0.4373393
Тернопільська	RS	Ризики та безпека життя	2016	0.70946111
Харківська	RS	Ризики та безпека життя	2016	0.31746312
Херсонська	RS	Ризики та безпека життя	2016	0.68351967
Хмельницька	RS	Ризики та безпека життя	2016	0.65453896
Черкаська	RS	Ризики та безпека життя	2016	0.61133143
Чернівецька	RS	Ризики та безпека життя	2016	0.47042901
Чернігівська	RS	Ризики та безпека життя	2016	0.63494596
Луганська	Iql	Індекс якості життя	2016	0.52623118
Львівська	Iql	Індекс якості життя	2016	0.48198686
Миколаївська	Iql	Індекс якості життя	2016	0.59380806
Одеська	Iql	Індекс якості життя	2016	0.47793123
Полтавська	Iql	Індекс якості життя	2016	0.50824192
Рівненська	Iql	Індекс якості життя	2016	0.47990594
Сумська	Iql	Індекс якості життя	2016	0.46148221

Тернопільська	Iql	Індекс якості життя	2016	0.52119344
Харківська	Iql	Індекс якості життя	2016	0.49060616
Херсонська	Iql	Індекс якості життя	2016	0.56388994
Хмельницька	Iql	Індекс якості життя	2016	0.51129645
Черкаська	Iql	Індекс якості життя	2016	0.51113766
Чернівецька	Iql	Індекс якості життя	2016	0.48387294
Чернігівська	Iql	Індекс якості життя	2016	0.5420351
Вінницька	Is	Індекс соціально-інституціонального виміру	2016	0.58911717
Волинська	Is	Індекс соціально-інституціонального виміру	2016	0.57061128
Дніпропетровська	Is	Індекс соціально-інституціонального виміру	2016	0.50369302
Донецька	Is	Індекс соціально-інституціонального виміру	2016	0.48798441
Житомирська	Is	Індекс соціально-інституціонального виміру	2016	0.55067809
Закарпатська	Is	Індекс соціально-інституціонального виміру	2016	0.59805137
Запорізька	Is	Індекс соціально-інституціонального виміру	2016	0.55138919
Івано-Франківська	Is	Індекс соціально-інституціонального виміру	2016	0.56773783
Київська	Is	Індекс соціально-інституціонального виміру	2016	0.53556469
Київ	Is	Індекс соціально-інституціонального виміру	2016	0.58425241
Кіровоградська	Is	Індекс соціально-інституціонального виміру	2016	0.50453812
Луганська	Is	Індекс соціально-інституціонального виміру	2016	0.54344018
Львівська	Is	Індекс соціально-інституціонального виміру	2016	0.56618518
Миколаївська	Is	Індекс соціально-інституціонального виміру	2016	0.54807655
Одеська	Is	Індекс соціально-інституціонального виміру	2016	0.51395661
Полтавська	Is	Індекс соціально-інституціонального виміру	2016	0.53858792
Рівненська	Is	Індекс соціально-інституціонального виміру	2016	0.54524642
Сумська	Is	Індекс соціально-інституціонального виміру	2016	0.50178222
Тернопільська	Is	Індекс соціально-інституціонального виміру	2016	0.58440805
Харківська	Is	Індекс соціально-інституціонального виміру	2016	0.5424741

Херсонська	Is	Індекс соціально-інституціонального виміру	2016	0.56109154
Хмельницька	Is	Індекс соціально-інституціонального виміру	2016	0.58918016
Черкаська	Is	Індекс соціально-інституціонального виміру	2016	0.56333203
Чернівецька	Is	Індекс соціально-інституціонального виміру	2016	0.55092385
Чернігівська	Is	Індекс соціально-інституціонального виміру	2016	0.53326288
Вінницька	Cql	Компонента якості життя	2016	0.80385719
Волинська	Cql	Компонента якості життя	2016	0.80540309
Дніпропетровська	Cql	Компонента якості життя	2016	0.76437239
Донецька	Cql	Компонента якості життя	2016	0.61095673
Житомирська	Cql	Компонента якості життя	2016	0.81591638
Закарпатська	Cql	Компонента якості життя	2016	0.86876808
Запорізька	Cql	Компонента якості життя	2016	0.71205486
Івано-Франківська	Cql	Компонента якості життя	2016	0.77583479
Київська	Cql	Компонента якості життя	2016	0.81798355
Київ	Cql	Компонента якості життя	2016	1.13716109
Кіровоградська	Cql	Компонента якості життя	2016	0.74969061
Луганська	Cql	Компонента якості життя	2016	0.70113397
Львівська	Cql	Компонента якості життя	2016	0.82914066
Миколаївська	Cql	Компонента якості життя	2016	0.76672162
Одеська	Cql	Компонента якості життя	2016	0.74826157
Полтавська	Cql	Компонента якості життя	2016	0.7929527
Рівненська	Cql	Компонента якості життя	2016	0.80441061
Сумська	Cql	Компонента якості життя	2016	0.74531798
Тернопільська	Cql	Компонента якості життя	2016	0.81232411
Харківська	Cql	Компонента якості життя	2016	0.85612153
Херсонська	Cql	Компонента якості життя	2016	0.75551243
Хмельницька	Cql	Компонента якості життя	2016	0.84475775
Черкаська	Cql	Компонента якості життя	2016	0.77685301
Чернівецька	Cql	Компонента якості життя	2016	0.82705141
Чернігівська	Cql	Компонента якості життя	2016	0.77625652
Вінницька	Se	Екологічна безпека	2016	0.52334572
Волинська	Se	Екологічна безпека	2016	0.55415466
Дніпропетровська	Se	Екологічна безпека	2016	0.35666715
Донецька	Se	Екологічна безпека	2016	0.26061865
Житомирська	Se	Екологічна безпека	2016	0.55069653
Закарпатська	Se	Екологічна безпека	2016	0.72046713
Запорізька	Se	Екологічна безпека	2016	0.45555509
Івано-Франківська	Se	Екологічна безпека	2016	0.58435849
Київська	Se	Екологічна безпека	2016	0.52536572
Київ	Se	Екологічна безпека	2016	0.29317792
Кіровоградська	Se	Екологічна безпека	2016	0.51188056
Луганська	Se	Екологічна безпека	2016	0.48347677

Львівська	Se	Екологічна безпека	2016	0.59664479
Миколаївська	Se	Екологічна безпека	2016	0.49765704
Одеська	Se	Екологічна безпека	2016	0.48792887
Полтавська	Se	Екологічна безпека	2016	0.52040308
Рівненська	Se	Екологічна безпека	2016	0.55796561
Сумська	Se	Екологічна безпека	2016	0.54808348
Тернопільська	Se	Екологічна безпека	2016	0.56557474
Харківська	Se	Екологічна безпека	2016	0.48777161
Херсонська	Se	Екологічна безпека	2016	0.4980274
Хмельницька	Se	Екологічна безпека	2016	0.54504917
Черкаська	Se	Екологічна безпека	2016	0.50361768
Чернівецька	Se	Екологічна безпека	2016	0.5796913
Чернігівська	Se	Екологічна безпека	2016	0.5510263
Івано-Франківська	SP	Сфера послуг	2016	0.00997694
Київська	SP	Сфера послуг	2016	0.05145735
Київ	SP	Сфера послуг	2016	0.333
Кіровоградська	SP	Сфера послуг	2016	0.010182
Луганська	SP	Сфера послуг	2016	0.00662602
Львівська	SP	Сфера послуг	2016	0.02944299
Миколаївська	SP	Сфера послуг	2016	0.01628039
Одеська	SP	Сфера послуг	2016	0.03826855
Полтавська	SP	Сфера послуг	2016	0.02816537
Рівненська	SP	Сфера послуг	2016	0.00737751
Сумська	SP	Сфера послуг	2016	0.01008265
Тернопільська	SP	Сфера послуг	2016	0.00692989
Харківська	SP	Сфера послуг	2016	0.03960878
Херсонська	SP	Сфера послуг	2016	0.00808676
Хмельницька	SP	Сфера послуг	2016	0.01002363
Черкаська	SP	Сфера послуг	2016	0.01682138
Чернівецька	SP	Сфера послуг	2016	0.00385783
Чернігівська	SP	Сфера послуг	2016	0.01077881
Вінницька	Ibr	Базові потреби	2016	0.34867406
Волинська	Ibr	Базові потреби	2016	0.12717817
Дніпропетровська	Ibr	Базові потреби	2016	0.50458795
Донецька	Ibr	Базові потреби	2016	0.2627612
Житомирська	Ibr	Базові потреби	2016	0.16709389
Закарпатська	Ibr	Базові потреби	2016	0.08292117
Запорізька	Ibr	Базові потреби	2016	0.24929498
Івано-Франківська	Ibr	Базові потреби	2016	0.12541055
Київська	Ibr	Базові потреби	2016	0.34511163
Київ	Ibr	Базові потреби	2016	0.83061744
Кіровоградська	Ibr	Базові потреби	2016	0.19943973
Луганська	Ibr	Базові потреби	2016	0.09341682
Львівська	Ibr	Базові потреби	2016	0.23157398
Миколаївська	Ibr	Базові потреби	2016	0.18621177
Одеська	Ibr	Базові потреби	2016	0.27455736

Полтавська	Ibr	Базові потреби	2016	0.34503735
Рівненська	Ibr	Базові потреби	2016	0.12614265
Сумська	Ibr	Базові потреби	2016	0.17960467
Тернопільська	Ibr	Базові потреби	2016	0.14044834
Харківська	Ibr	Базові потреби	2016	0.35826827
Херсонська	Ibr	Базові потреби	2016	0.1823681
Хмельницька	Ibr	Базові потреби	2016	0.20963775
Черкаська	Ibr	Базові потреби	2016	0.25877455
Чернівецька	Ibr	Базові потреби	2016	0.07487684
Чернігівська	Ibr	Базові потреби	2016	0.17719498
Вінницька	PA	Підприємницька активність	2016	0.06357295
Волинська	PA	Підприємницька активність	2016	0.04388792
Дніпропетровська	PA	Підприємницька активність	2016	0.14618551
Донецька	PA	Підприємницька активність	2016	0.10132947
Житомирська	PA	Підприємницька активність	2016	0.05034203
Закарпатська	PA	Підприємницька активність	2016	0.04517874
Запорізька	PA	Підприємницька активність	2016	0.08196715
Івано-Франківська	PA	Підприємницька активність	2016	0.05356908
Київська	PA	Підприємницька активність	2016	0.10810628
Київ	PA	Підприємницька активність	2016	0.334
Кіровоградська	PA	Підприємницька активність	2016	0.04537237
Луганська	PA	Підприємницька активність	2016	0.04582415
Львівська	PA	Підприємницька активність	2016	0.11165604
Миколаївська	PA	Підприємницька активність	2016	0.06292754
Одеська	PA	Підприємницька активність	2016	0.12069179
Полтавська	PA	Підприємницька активність	2016	0.06421836
Рівненська	PA	Підприємницька активність	2016	0.0425971
Сумська	PA	Підприємницька активність	2016	0.04130628
Тернопільська	PA	Підприємницька активність	2016	0.0425971
Харківська	PA	Підприємницька активність	2016	0.11391498
Херсонська	PA	Підприємницька активність	2016	0.04937391
Хмельницька	PA	Підприємницька активність	2016	0.05421449
Черкаська	PA	Підприємницька активність	2016	0.05744155
Чернівецька	PA	Підприємницька активність	2016	0.03194783
Чернігівська	PA	Підприємницька активність	2016	0.04162899
Вінницька	SR	Споживчий ринок	2016	0.06728125
Волинська	SR	Споживчий ринок	2016	0.05021172
Дніпропетровська	SR	Споживчий ринок	2016	0.19559025
Донецька	SR	Споживчий ринок	2016	0.0677466
Житомирська	SR	Споживчий ринок	2016	0.0553857
Закарпатська	SR	Споживчий ринок	2016	0.0484102
Запорізька	SR	Споживчий ринок	2016	0.10019646
Івано-Франківська	SR	Споживчий ринок	2016	0.06834162
Київська	SR	Споживчий ринок	2016	0.10691985
Київ	SR	Споживчий ринок	2016	0.333
Кіровоградська	SR	Споживчий ринок	2016	0.04439244

Луганська	SR	Споживчий ринок	2016	0.01637016
Львівська	SR	Споживчий ринок	2016	0.12304042
Миколаївська	SR	Споживчий ринок	2016	0.05861057
Одеська	SR	Споживчий ринок	2016	0.15202216
Полтавська	SR	Споживчий ринок	2016	0.06283921
Рівненська	SR	Споживчий ринок	2016	0.04769834
Сумська	SR	Споживчий ринок	2016	0.04664376
Тернопільська	SR	Споживчий ринок	2016	0.04056477
Харківська	SR	Споживчий ринок	2016	0.17324093
Херсонська	SR	Споживчий ринок	2016	0.05355035
Хмельницька	SR	Споживчий ринок	2016	0.05253918
Черкаська	SR	Споживчий ринок	2016	0.05463966
Чернівецька	SR	Споживчий ринок	2016	0.03382094
Чернігівська	SR	Споживчий ринок	2016	0.04357659
Вінницька	ZED	Зовнішньоекономічна діяльність	2016	0.02147862
Волинська	ZED	Зовнішньоекономічна діяльність	2016	0.02025009
Дніпропетровська	ZED	Зовнішньоекономічна діяльність	2016	0.13736726
Донецька	ZED	Зовнішньоекономічна діяльність	2016	0.07842838
Житомирська	ZED	Зовнішньоекономічна діяльність	2016	0.01212627
Закарпатська	ZED	Зовнішньоекономічна діяльність	2016	0.03377464
Запорізька	ZED	Зовнішньоекономічна діяльність	2016	0.05257043
Івано-Франківська	ZED	Зовнішньоекономічна діяльність	2016	0.01449564
Київська	ZED	Зовнішньоекономічна діяльність	2016	0.05735039
Київ	ZED	Зовнішньоекономічна діяльність	2016	0.333
Кіровоградська	ZED	Зовнішньоекономічна діяльність	2016	0.0097627
Луганська	ZED	Зовнішньоекономічна діяльність	2016	0.01153856
Львівська	ZED	Зовнішньоекономічна діяльність	2016	0.04318831
Миколаївська	ZED	Зовнішньоекономічна діяльність	2016	0.04358198
Одеська	ZED	Зовнішньоекономічна діяльність	2016	0.0518091
Полтавська	ZED	Зовнішньоекономічна діяльність	2016	0.03387042
Рівненська	ZED	Зовнішньоекономічна діяльність	2016	0.00827892
Сумська	ZED	Зовнішньоекономічна діяльність	2016	0.0136793
Тернопільська	ZED	Зовнішньоекономічна діяльність	2016	0.00815069
Харківська	ZED	Зовнішньоекономічна діяльність	2016	0.03469043
Херсонська	ZED	Зовнішньоекономічна діяльність	2016	0.00631003
Хмельницька	ZED	Зовнішньоекономічна діяльність	2016	0.00861251
Черкаська	ZED	Зовнішньоекономічна діяльність	2016	0.01124978
Чернівецька	ZED	Зовнішньоекономічна діяльність	2016	0.00324897
Чернігівська	ZED	Зовнішньоекономічна діяльність	2016	0.01170591
Вінницька	Ipd	Підприємницька діяльність	2016	0.15233283
Волинська	Ipd	Підприємницька діяльність	2016	0.11434973
Дніпропетровська	Ipd	Підприємницька діяльність	2016	0.47914301
Донецька	Ipd	Підприємницька діяльність	2016	0.24750445
Житомирська	Ipd	Підприємницька діяльність	2016	0.117854
Закарпатська	Ipd	Підприємницька діяльність	2016	0.12736358
Хмельницька	So	Безпека (інше)	2016	0.5

Черкаська	So	Безпека (інше)	2016	0.5
Чернівецька	So	Безпека (інше)	2016	0.5
Чернігівська	So	Безпека (інше)	2016	0.5
Вінницька	Csl	Компонента безпеки життя	2016	1.23164475
Волинська	Csl	Компонента безпеки життя	2016	1.18393767
Дніпропетровська	Csl	Компонента безпеки життя	2016	0.9690492
Донецька	Csl	Компонента безпеки життя	2016	1.18805766
Житомирська	Csl	Компонента безпеки життя	2016	1.06980425
Закарпатська	Csl	Компонента безпеки життя	2016	1.25520274
Запорізька	Csl	Компонента безпеки життя	2016	1.0422702
Івано-Франківська	Csl	Компонента безпеки життя	2016	1.40958103
Київська	Csl	Компонента безпеки життя	2016	1.0143771
Київ	Csl	Компонента безпеки життя	2016	1.37755739
Кіровоградська	Csl	Компонента безпеки життя	2016	1.03909019
Луганська	Csl	Компонента безпеки життя	2016	1.24860986
Львівська	Csl	Компонента безпеки життя	2016	1.25092073
Миколаївська	Csl	Компонента безпеки життя	2016	1.14540082
Одеська	Csl	Компонента безпеки життя	2016	0.98353627
Полтавська	Csl	Компонента безпеки життя	2016	1.12268195
Рівненська	Csl	Компонента безпеки життя	2016	1.17689593
Сумська	Csl	Компонента безпеки життя	2016	1.08792648
Тернопільська	Csl	Компонента безпеки життя	2016	1.30978036
Харківська	Csl	Компонента безпеки життя	2016	1.07807323
Херсонська	Csl	Компонента безпеки життя	2016	1.01497017
Хмельницька	Csl	Компонента безпеки життя	2016	1.16314182
Черкаська	Csl	Компонента безпеки життя	2016	1.1175885
Чернівецька	Csl	Компонента безпеки життя	2016	1.27231565
Чернігівська	Csl	Компонента безпеки життя	2016	1.08764465
Вінницька	Isd	Індекс сталого розвитку	2016	0.96726342
Волинська	Isd	Індекс сталого розвитку	2016	0.95713116
Дніпропетровська	Isd	Індекс сталого розвитку	2016	0.8737412
Донецька	Isd	Індекс сталого розвитку	2016	0.80159748
Житомирська	Isd	Індекс сталого розвитку	2016	0.94025006
Закарпатська	Isd	Індекс сталого розвитку	2016	1.02730099
Запорізька	Isd	Індекс сталого розвитку	2016	0.84514958
Івано-Франківська	Isd	Індекс сталого розвитку	2016	0.99045208
Київська	Isd	Індекс сталого розвитку	2016	0.93027314
Київ	Isd	Індекс сталого розвитку	2016	1.286539
Кіровоградська	Isd	Індекс сталого розвитку	2016	0.87637261
Луганська	Isd	Індекс сталого розвитку	2016	0.88827295
Львівська	Isd	Індекс сталого розвитку	2016	0.99298349
Миколаївська	Isd	Індекс сталого розвитку	2016	0.91551189
Одеська	Isd	Індекс сталого розвитку	2016	0.86281557
Полтавська	Isd	Індекс сталого розвитку	2016	0.93233208
Рівненська	Isd	Індекс сталого розвитку	2016	0.9546364
Сумська	Isd	Індекс сталого розвитку	2016	0.88391731

Тернопільська	Isd	Індекс сталого розвитку	2016	0.9935633
Харківська	Isd	Індекс сталого розвитку	2016	0.97707891
Херсонська	Isd	Індекс сталого розвитку	2016	0.87597927
Хмельницька	Isd	Індекс сталого розвитку	2016	0.98576338
Черкаська	Isd	Індекс сталого розвитку	2016	0.91749263
Чернівецька	Isd	Індекс сталого розвитку	2016	0.99642134
Чернігівська	Isd	Індекс сталого розвитку	2016	0.91009237
Вінницька	G	Ступінь гармонізації	2016	0.61377405
Волинська	G	Ступінь гармонізації	2016	0.49373729
Дніпропетровська	G	Ступінь гармонізації	2016	0.83703986
Донецька	G	Ступінь гармонізації	2016	0.63671884
Житомирська	G	Ступінь гармонізації	2016	0.54791961
Закарпатська	G	Ступінь гармонізації	2016	0.52289016
Запорізька	G	Ступінь гармонізації	2016	0.65109931
Івано-Франківська	G	Ступінь гармонізації	2016	0.50389578
Київська	G	Ступінь гармонізації	2016	0.74431922
Київ	G	Ступінь гармонізації	2016	0.72671134
Кіровоградська	G	Ступінь гармонізації	2016	0.55293034
Луганська	G	Ступінь гармонізації	2016	0.5369027
Львівська	G	Ступінь гармонізації	2016	0.66766128
Миколаївська	G	Ступінь гармонізації	2016	0.59859568
Одеська	G	Ступінь гармонізації	2016	0.71788603
Полтавська	G	Ступінь гармонізації	2016	0.64678963
Рівненська	G	Ступінь гармонізації	2016	0.51626586
Сумська	G	Ступінь гармонізації	2016	0.62368488
Тернопільська	G	Ступінь гармонізації	2016	0.48846988
Харківська	G	Ступінь гармонізації	2016	0.85870771
Херсонська	G	Ступінь гармонізації	2016	0.53333278
Хмельницька	G	Ступінь гармонізації	2016	0.51934775
Черкаська	G	Ступінь гармонізації	2016	0.57303889
Чернівецька	G	Ступінь гармонізації	2016	0.51538078
Чернігівська	G	Ступінь гармонізації	2016	0.53315292
Вінницька	VNP	Валовий національний продукт	2016	0.0441378
Волинська	VNP	Валовий національний продукт	2016	0.02336087
Дніпропетровська	VNP	Валовий національний продукт	2016	0.15865308
Донецька	VNP	Валовий національний продукт	2016	0.08478857
Житомирська	VNP	Валовий національний продукт	2016	0.02832749
Закарпатська	VNP	Валовий національний продукт	2016	0.02134385
Запорізька	VNP	Валовий національний продукт	2016	0.0656571
Івано-Франківська	VNP	Валовий національний продукт	2016	0.03380426
Київська	VNP	Валовий національний продукт	2016	0.07669247
Київ	VNP	Валовий національний продукт	2016	0.333
Кіровоградська	VNP	Валовий національний продукт	2016	0.0283437
Луганська	VNP	Валовий національний продукт	2016	0.01758184
Львівська	VNP	Валовий національний продукт	2016	0.06980689
Миколаївська	VNP	Валовий національний продукт	2016	0.03553008

Одеська	VNP	Валовий національний продукт	2016	0.0735453
Полтавська	VNP	Валовий національний продукт	2016	0.07067459
Рівненська	VNP	Валовий національний продукт	2016	0.0259883
Сумська	VNP	Валовий національний продукт	2016	0.03064381
Тернопільська	VNP	Валовий національний продукт	2016	0.0196512
Харківська	VNP	Валовий національний продукт	2016	0.09203613
Херсонська	VNP	Валовий національний продукт	2016	0.02374938
Хмельницька	VNP	Валовий національний продукт	2016	0.03029069
Черкаська	VNP	Валовий національний продукт	2016	0.03748222
Чернівецька	VNP	Валовий національний продукт	2016	0.0136429
Чернігівська	VNP	Валовий національний продукт	2016	0.02725189
Вінницька	VSF	Виробнича сфера	2016	0.28724588
Волинська	VSF	Виробнича сфера	2016	0.08937559
Дніпропетровська	VSF	Виробнича сфера	2016	0.25399078
Донецька	VSF	Виробнича сфера	2016	0.13386918
Житомирська	VSF	Виробнича сфера	2016	0.12792162
Закарпатська	VSF	Виробнича сфера	2016	0.05473457
Запорізька	VSF	Виробнича сфера	2016	0.15419223
Івано-Франківська	VSF	Виробнича сфера	2016	0.08162935
Київська	VSF	Виробнича сфера	2016	0.21696181
Київ	VSF	Виробнича сфера	2016	0.16461744
Кіровоградська	VSF	Виробнича сфера	2016	0.16091403
Луганська	VSF	Виробнича сфера	2016	0.06920896
Львівська	VSF	Виробнича сфера	2016	0.1323241
Миколаївська	VSF	Виробнича сфера	2016	0.1344013
Одеська	VSF	Виробнича сфера	2016	0.1627435
Полтавська	VSF	Виробнича сфера	2016	0.24619738
Рівненська	VSF	Виробнича сфера	2016	0.09277684
Сумська	VSF	Виробнича сфера	2016	0.13887821
Тернопільська	VSF	Виробнича сфера	2016	0.11386725
Харківська	VSF	Виробнича сфера	2016	0.22662337
Херсонська	VSF	Виробнича сфера	2016	0.15053196
Хмельницька	VSF	Виробнича сфера	2016	0.16932344
Черкаська	VSF	Виробнича сфера	2016	0.20447095
Чернівецька	VSF	Виробнича сфера	2016	0.05737611
Чернігівська	VSF	Виробнича сфера	2016	0.13916428
Вінницька	SP	Сфера послуг	2016	0.01729039
Волинська	SP	Сфера послуг	2016	0.01444171
Дніпропетровська	SP	Сфера послуг	2016	0.09194409
Донецька	SP	Сфера послуг	2016	0.04410345
Житомирська	SP	Сфера послуг	2016	0.01084479
Закарпатська	SP	Сфера послуг	2016	0.00684276
Запорізька	SP	Сфера послуг	2016	0.02944565
Вінницька	Ss	Соціальна безпека	2016	1.07909601
Волинська	Ss	Соціальна безпека	2016	0.96506026
Дніпропетровська	Ss	Соціальна безпека	2016	0.67259834

Донецька	Ss	Соціальна безпека	2016	0.92059036
Житомирська	Ss	Соціальна безпека	2016	0.77924139
Закарпатська	Ss	Соціальна безпека	2016	1.0595538
Запорізька	Ss	Соціальна безпека	2016	0.73551929
Івано-Франківська	Ss	Соціальна безпека	2016	1.3043985
Київська	Ss	Соціальна безпека	2016	0.76042483
Київ	Ss	Соціальна безпека	2016	1.11194083
Кіровоградська	Ss	Соціальна безпека	2016	0.63311328
Луганська	Ss	Соціальна безпека	2016	0.9218073
Львівська	Ss	Соціальна безпека	2016	1.10575971
Миколаївська	Ss	Соціальна безпека	2016	0.96922107
Одеська	Ss	Соціальна безпека	2016	0.73320165
Полтавська	Ss	Соціальна безпека	2016	0.76159961
Рівненська	Ss	Соціальна безпека	2016	0.98458648
Сумська	Ss	Соціальна безпека	2016	0.87211976
Тернопільська	Ss	Соціальна безпека	2016	1.14206771
Харківська	Ss	Соціальна безпека	2016	0.89405928
Херсонська	Ss	Соціальна безпека	2016	0.69657706
Хмельницька	Ss	Соціальна безпека	2016	0.98968158
Черкаська	Ss	Соціальна безпека	2016	0.91844158
Чернівецька	Ss	Соціальна безпека	2016	1.14874083
Чернігівська	Ss	Соціальна безпека	2016	0.80442489
Вінницька	TH	Техногенна небезпека	2016	0.5
Волинська	TH	Техногенна небезпека	2016	0.5
Дніпропетровська	TH	Техногенна небезпека	2016	0.5
Донецька	TH	Техногенна небезпека	2016	0.5
Житомирська	TH	Техногенна небезпека	2016	0.5
Закарпатська	TH	Техногенна небезпека	2016	0.5
Запорізька	TH	Техногенна небезпека	2016	0.5
Івано-Франківська	TH	Техногенна небезпека	2016	0.5
Київська	TH	Техногенна небезпека	2016	0.5
Київ	TH	Техногенна небезпека	2016	0.5
Кіровоградська	TH	Техногенна небезпека	2016	0.5
Луганська	TH	Техногенна небезпека	2016	0.5
Львівська	TH	Техногенна небезпека	2016	0.5
Миколаївська	TH	Техногенна небезпека	2016	0.5
Одеська	TH	Техногенна небезпека	2016	0.5
Полтавська	TH	Техногенна небезпека	2016	0.5
Рівненська	TH	Техногенна небезпека	2016	0.5
Сумська	TH	Техногенна небезпека	2016	0.5
Тернопільська	TH	Техногенна небезпека	2016	0.5
Харківська	TH	Техногенна небезпека	2016	0.5
Херсонська	TH	Техногенна небезпека	2016	0.5
Хмельницька	TH	Техногенна небезпека	2016	0.5
Черкаська	TH	Техногенна небезпека	2016	0.5
Чернівецька	TH	Техногенна небезпека	2016	0.5

Чернігівська	TH	Техногенна небезпека	2016	0.5
Вінницька	So	Безпека (інше)	2016	0.5
Волинська	So	Безпека (інше)	2016	0.5
Дніпропетровська	So	Безпека (інше)	2016	0.5
Донецька	So	Безпека (інше)	2016	0.5
Житомирська	So	Безпека (інше)	2016	0.5
Закарпатська	So	Безпека (інше)	2016	0.5
Запорізька	So	Безпека (інше)	2016	0.5
Івано-Франківська	So	Безпека (інше)	2016	0.5
Київська	So	Безпека (інше)	2016	0.5
Київ	So	Безпека (інше)	2016	0.5
Кіровоградська	So	Безпека (інше)	2016	0.5
Луганська	So	Безпека (інше)	2016	0.5
Львівська	So	Безпека (інше)	2016	0.5
Миколаївська	So	Безпека (інше)	2016	0.5
Одеська	So	Безпека (інше)	2016	0.5
Полтавська	So	Безпека (інше)	2016	0.5
Рівненська	So	Безпека (інше)	2016	0.5
Сумська	So	Безпека (інше)	2016	0.5
Тернопільська	So	Безпека (інше)	2016	0.5
Харківська	So	Безпека (інше)	2016	0.5
Херсонська	So	Безпека (інше)	2016	0.5
Вінницька	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.4640126
Волинська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.42205786
Дніпропетровська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.80645213
Донецька	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.5
Житомирська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.5514026
Закарпатська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.28515944
Запорізька	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.67383126
Івано-Франківська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.27294092
Київська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.59328889
Київ	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.2748576
Кіровоградська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.62914172
Луганська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.5
Львівська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.34874723
Миколаївська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.50638326
Одеська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.61636595
Полтавська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.62589611
Рівненська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.31379376
Сумська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.57735553
Тернопільська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.34331661
Харківська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.53521709
Херсонська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.59238474
Хмельницька	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.3974778
Черкаська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.54351555
Чернівецька	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.33235793

Чернігівська	HEALTH_DEC	Смертність населення за причинами смерті	2016	0.64855728
--------------	------------	--	------	------------

Додаток Б. Діагностика інтерполяційних геостатистичних моделей розподілу значень миш'яку у підземних водах території України

Для оцінки якості моделей проведемо перехресну перевірку для кожної з моделей. Кількість частин для розбиття візьмемо рівним 5. Обчислюватимемо норму вектора нев'язки та коефіцієнт детермінації.

Метод обернено зважених відстаней, ступінь 1:

```
set.seed(88)
spdf.gstat<-gstat(id = "AS1", formula = AS1 ~ 1, locations = spdf,
  set = list(idp = 1.0))
idw1cv<-gstat.cv(spdf.gstat,nfold =5)
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
pe <- idw1cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003682343
idw1cvdat<-as.data.frame(idw1cv)
ggplot(idw1cvdat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (IDW-1)", x ="long",y
="lat")
```

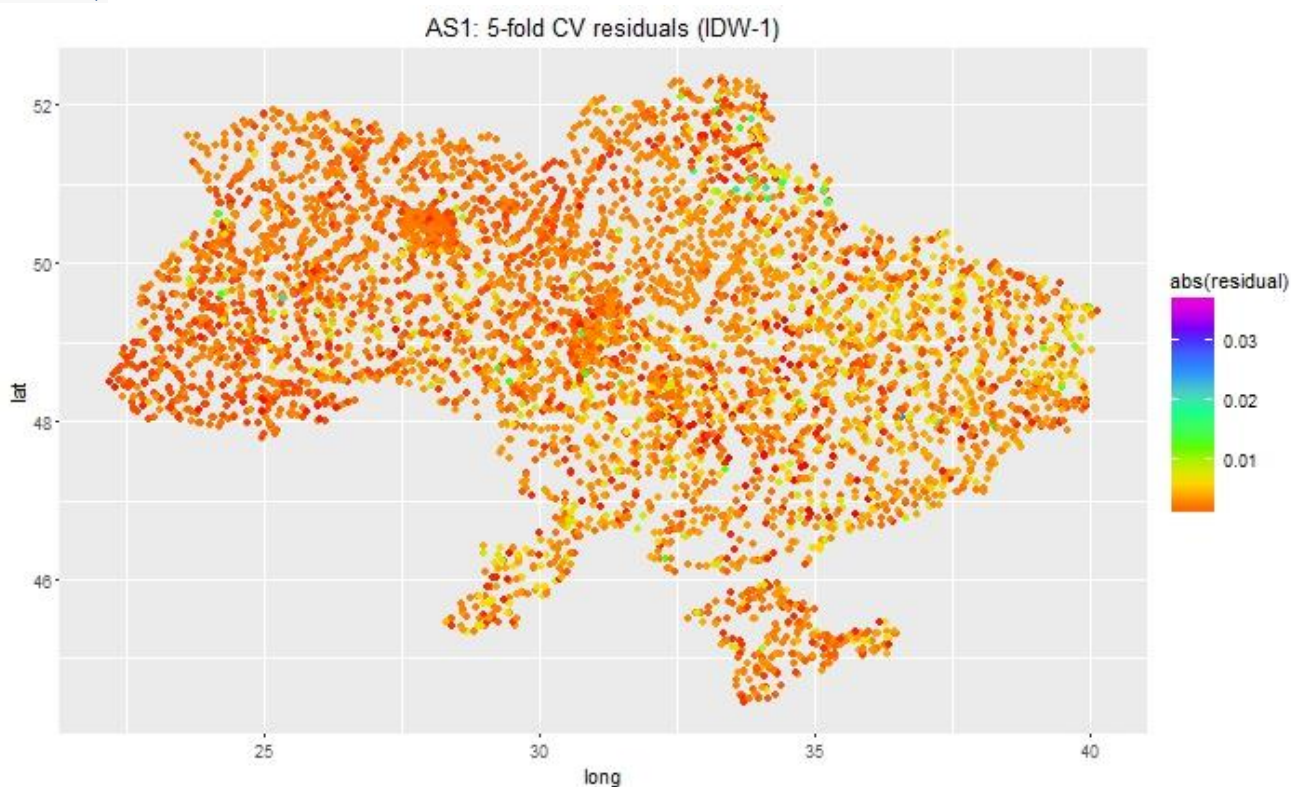


Рис. Б.1. Метод обернено зважених відстаней, ступінь 1

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(idw1cv@data$AS1.pred)/sd(spdf@data$AS1)
## [1] 0.1687547
```

Метод обернено зважених відстаней, ступінь 2:

```
spdf.gstat<-gstat(id = "AS1", formula = AS1 ~ 1, locations = spdf,
  set = list(idp = 2.0))
idw2cv<-gstat.cv(spdf.gstat,nfold =5)
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
pe <- idw2cv$residual
sqrt(mean(pe^2))
## [1] 0.003577271
idw2cvdat<-as.data.frame(idw2cv)
ggplot(idw2cvdat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (IDW-2)",x="long",y="lat")
```

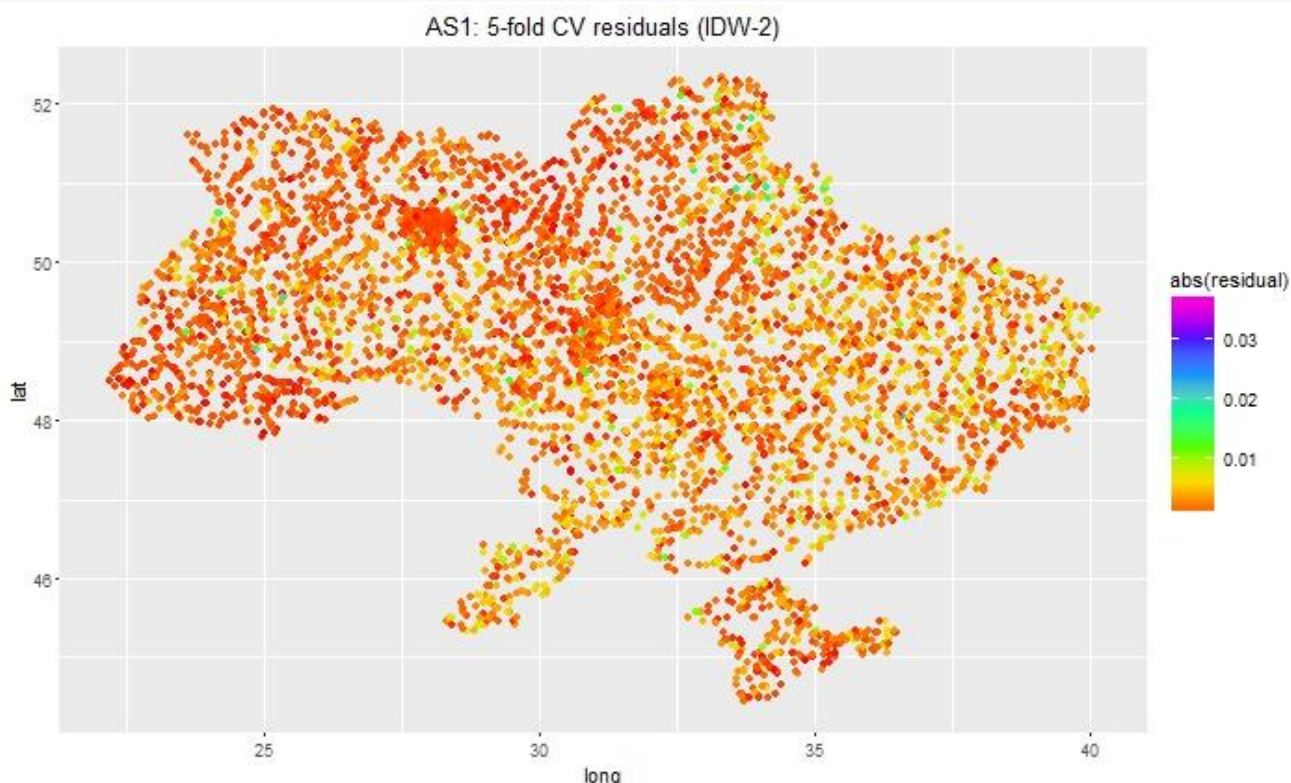


Рис. Б.2. Метод обернено зважених відстаней, ступінь 2:

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(idw2cv@data$AS1.pred)/sd(spdf@data$AS1)
## [1] 0.549687
```


Метод обернено зважених відстаней, ступінь 3:

```
spdf.gstat<-gstat(id = "AS1", formula = AS1 ~ 1, locations = spdf,
  set = list(idp = 3.0))
idw3cv<-gstat.cv(spdf.gstat,nfold =5)
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
## [inverse distance weighted interpolation]
pe <- idw3cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003934212
idw3cvdat<-as.data.frame(idw3cv)
ggplot(idw3cvdat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (IDW-3)",x="long",y="lat")
```

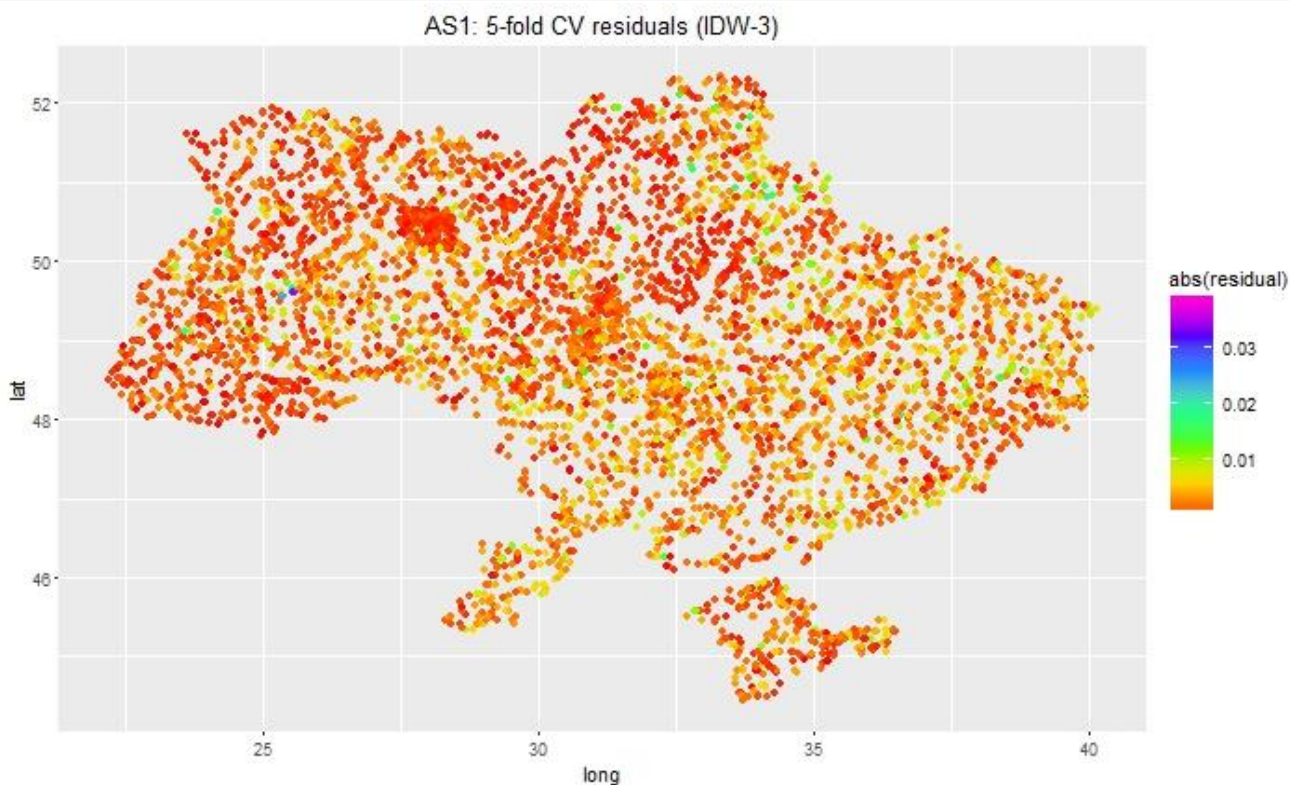


Рис. Б.3. Метод обернено зважених відстаней, ступінь 3:

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(idw3cv@data$AS1.pred)/sd(spdf@data$AS1)
## [1] 0.7788645
Крігінг, експоненційна модель:
model.variog.exp.cv<-vgm(psill=1.7e-05, model="Exp", nugget=1.0e-05, range=520)
krig.exp.cv <- krige.cv(AS1~1, spdf, model.variog.exp.cv, nfold=5)
pe<-krig.exp.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
```

```
sqrt(mean(pe^2))
## [1] 0.003400866
krigexpdat<-as.data.frame(krig.exp.cv)
ggplot(krigexpdat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-
Exp)",x="long",y="lat")
```

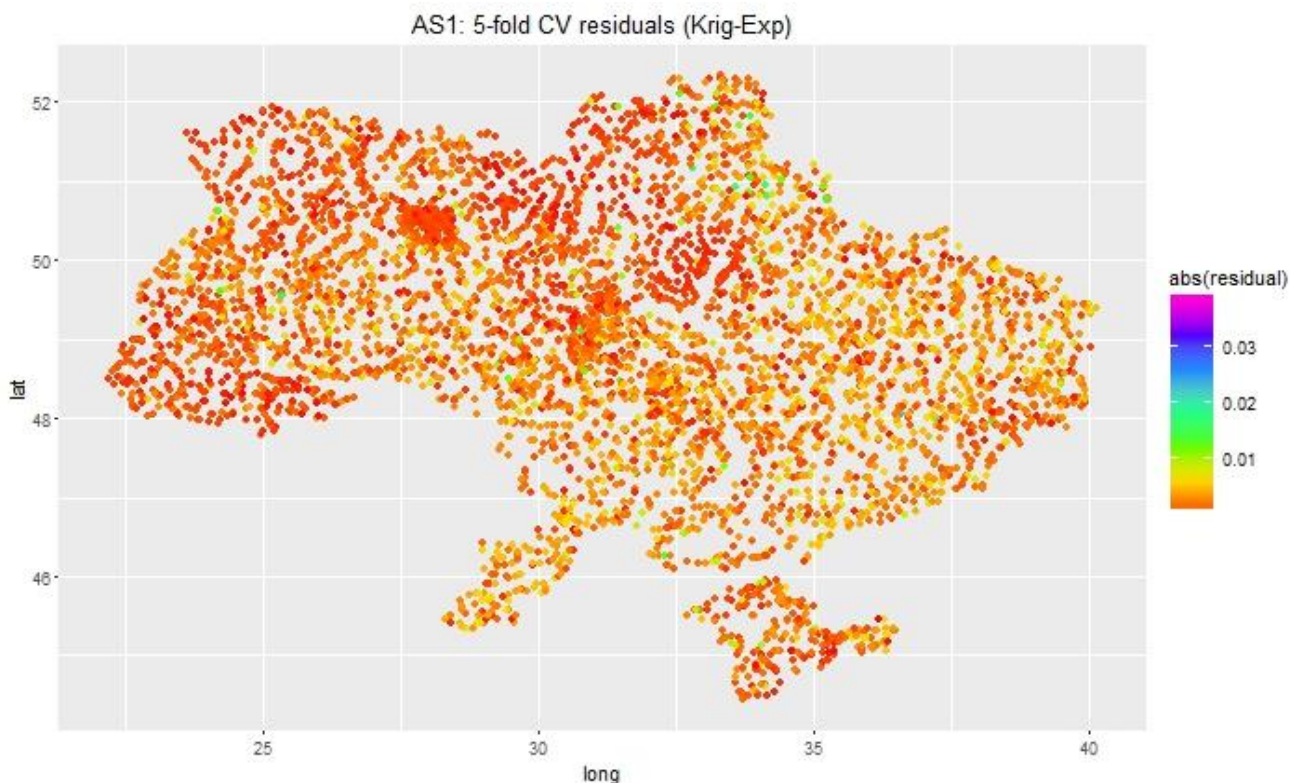


Рис. Б.4. Крігінг, експоненційна модель.

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.exp.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.4884747
Крігінг, гаусівська модель:
model.variog.gau.cv<-vgm(psill=1.7e-05, model="Gau", nugget=1.0e-05, range=520)
krig.gau.cv <- krige.cv(AS1~1, spdf, model.variog.gau.cv, nfold=5)
pe<-krig.gau.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003592975
kriggaudat<-as.data.frame(krig.gau.cv)
ggplot(kriggaudat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
```

```
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-
Gau)",x="long",y="lat")
```

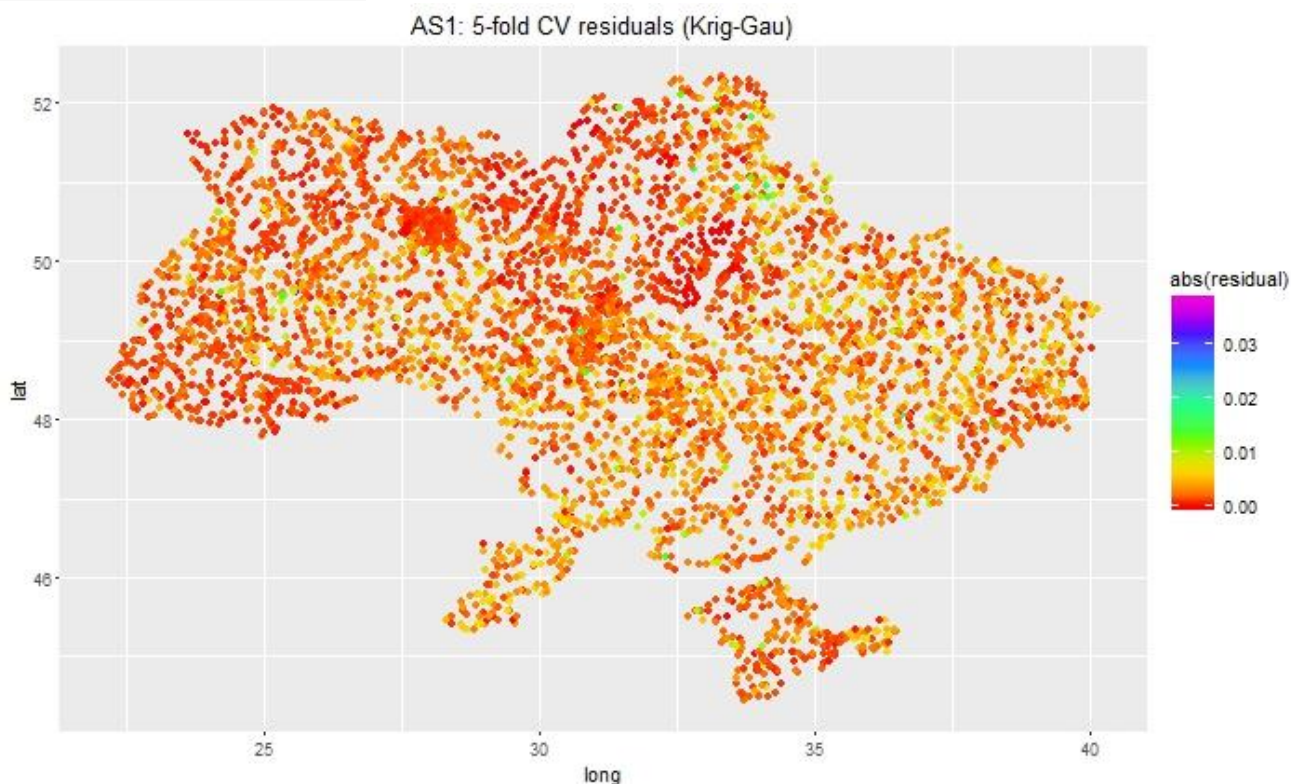


Рис. Б.5. Крігінг, гаусівська модель

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.gau.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.3921688
```

Крігінг, сферична модель:

```
model.variog.sph.cv<-vgm(psill=1.7e-05, model="Sph", nugget=1.0e-05, range=520)
krig.sph.cv <- krige.cv(AS1~1, spdf, model.variog.sph.cv, nfold=5)
pe<-krig.sph.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003395039
krigsphdat<-as.data.frame(krig.sph.cv)
ggplot(krigsphdat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
```



```
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-Sph)",x="long",y="lat")
```

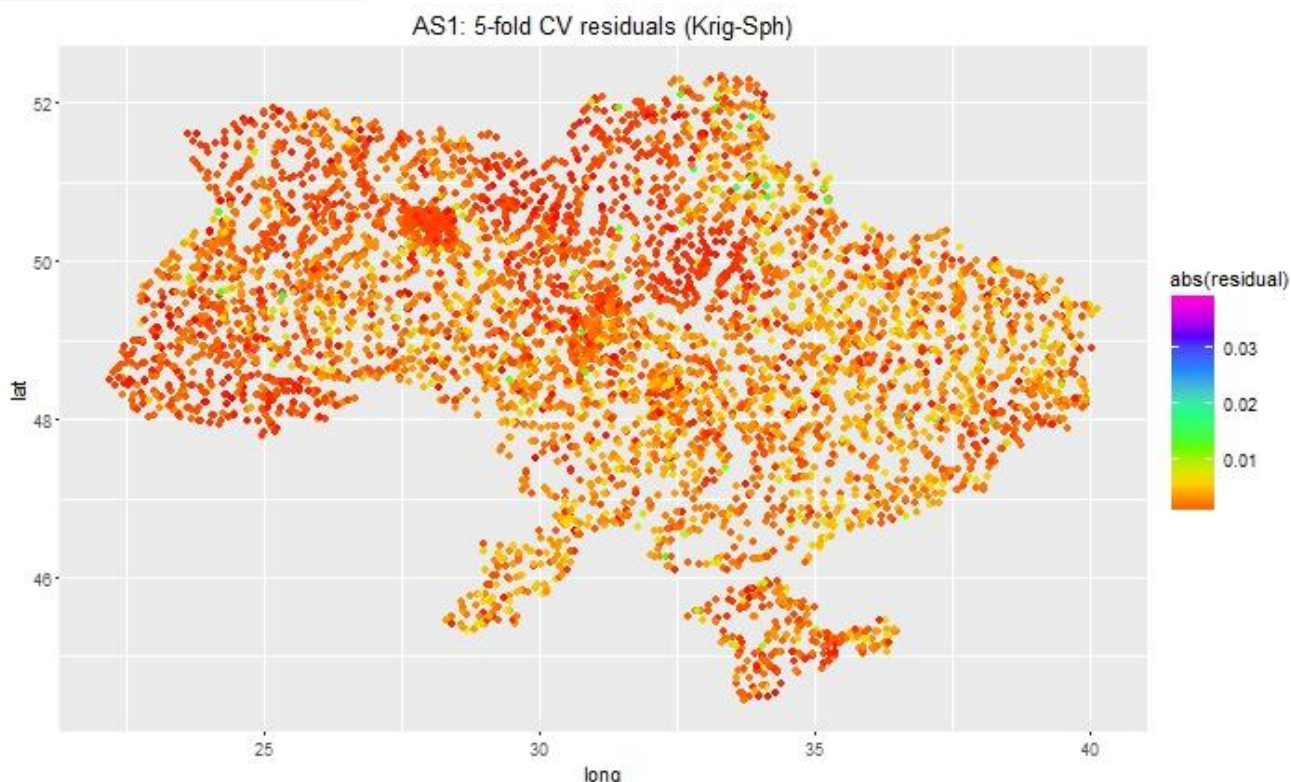


Рис. Б.6. Крігінг, сферична модель.

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.sph.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.502267
Ординарний крігінг:
model.variog.neig10.cv<-vgm(psill=1.7e-05, model="Exp", nugget=1.0e-05, range=520)
krig.neig10.cv <- krige.cv(AS1~1, spdf, model.variog.neig10.cv, nfold=5,nmax = 10)
pe<-krig.neig10.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003462349
krigneig10dat<-as.data.frame(krig.neig10.cv)
ggplot(krigneig10dat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-
Neigh10)",x="long",y="lat")
```

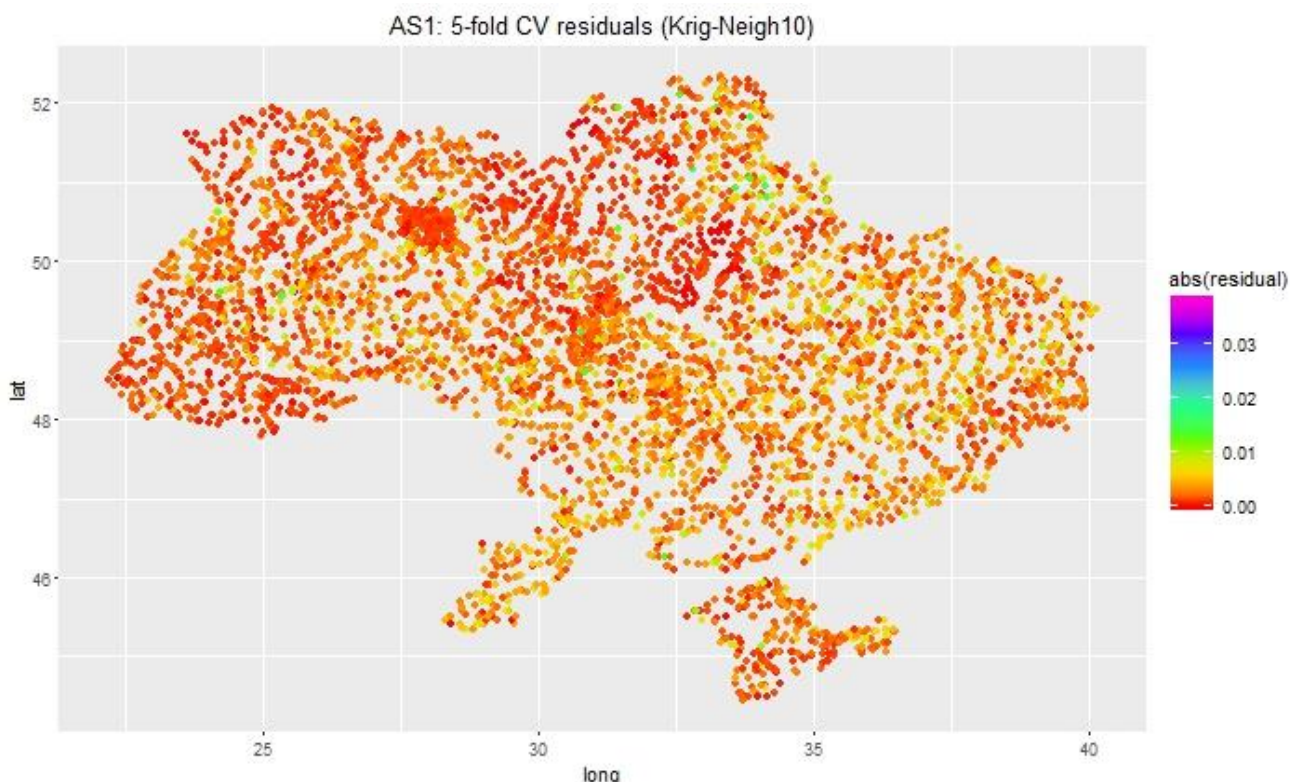


Рис. Б.7. Ординарний крігінг.

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.neig10.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.5831523
Кокрігінг:
model.variog.neig20.cv<-vgm(psill=1.7e-05, model="Exp", nugget=1.0e-05, range=520)
krig.neig20.cv <- krige.cv(AS1~1, spdf, model.variog.neig20.cv, nfold=5,nmax = 20)
pe<-krig.neig20.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003409049
krigneig20dat<-as.data.frame(krig.neig20.cv)
ggplot(krigneig20dat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-
Neigh20)",x="long",y="lat")
```

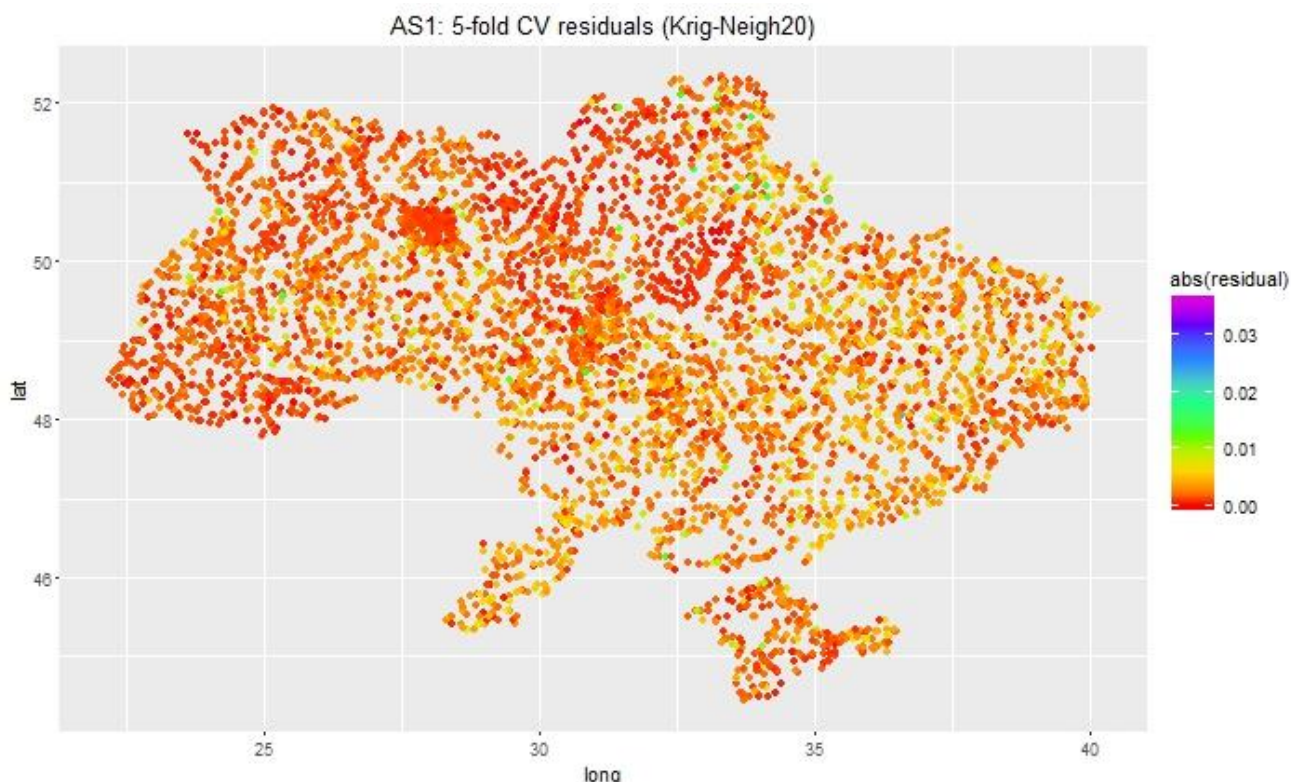


Рис. Б.8. Кокрігінг.

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.neig20.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.5306521
Індикаторний крігінг:
model.variog.neig30.cv<-vgm(psill=1.7e-05, model="Exp", nugget=1.0e-05, range=520)
krig.neig30.cv <- krige.cv(AS1~1, spdf, model.variog.neig30.cv, nfold=5,nmax = 30)
pe<-krig.neig30.cv$residual
print("Середньоквадратична помилка: ")
## [1] "Середньоквадратична помилка: "
sqrt(mean(pe^2))
## [1] 0.003413332
krigneig30dat<-as.data.frame(krig.neig30.cv)
ggplot(krigneig30dat,aes(coords.x1,coords.x2)) + geom_point(aes(colour =
abs(residual))) +
scale_colour_gradientn(colors = rainbow(7)) +
```



```
labs(fill = "Миш'як", title = "AS1: 5-fold CV residuals (Krig-Neigh30)", x="long", y="lat")
```

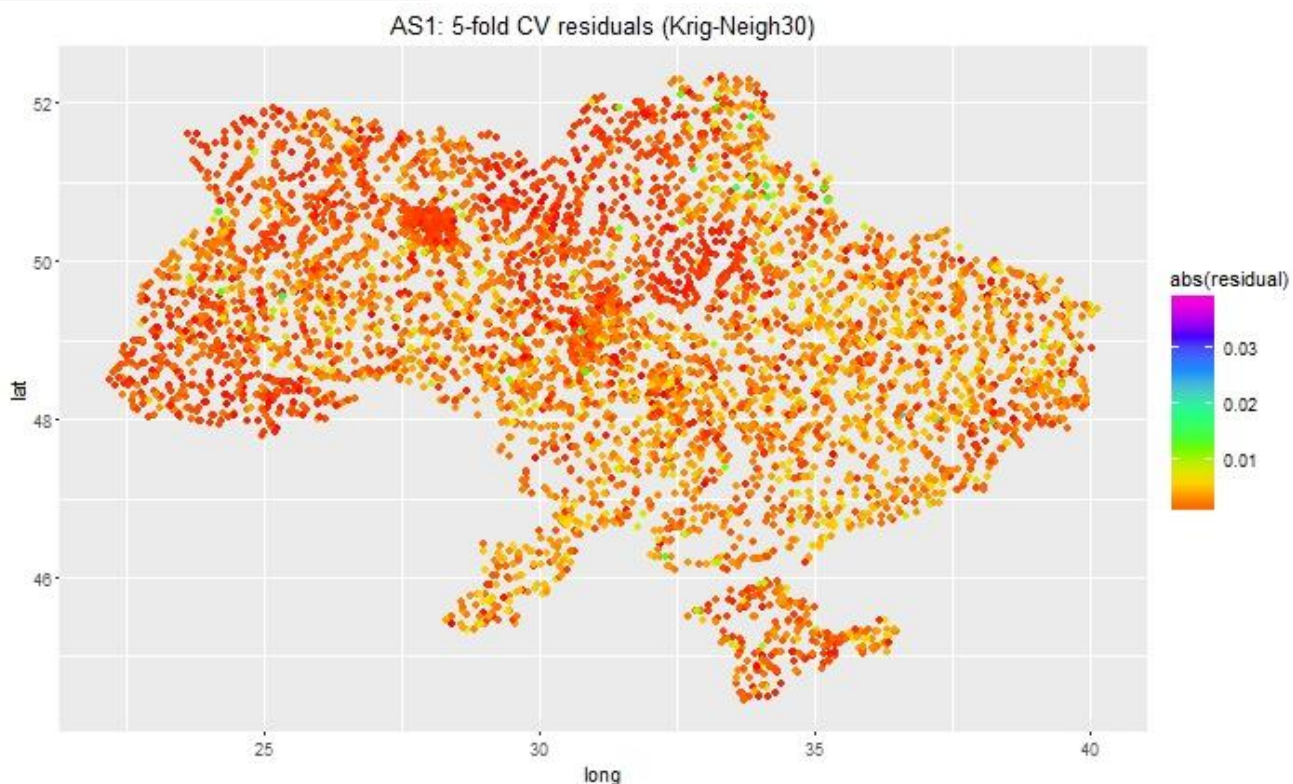


Рис. Б.9. Індикаторний крігінг.

```
print("Коефіцієнт детермінації: ")
## [1] "Коефіцієнт детермінації: "
sd(krig.neig30.cv@data$var1.pred)/sd(spdf@data$AS1)
## [1] 0.5115193
```

Додаток В. Копії актів впровадження



УКРАЇНА

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ННК «СВІТОВИЙ ЦЕНТР ДАНИХ З ГЕОІНФОРМАТИКИ ТА
СТАЛОГО РОЗВИТКУ»

03056, м. Київ, пр-т Перемоги, 37; тел. (+38 044) 204-94-55 тел./факс (+38 044) 236-10-43
<http://www.mmi.kpi.ua> e-mail: mmi@kpi.ua ЄДРПОУ 02070921

135/11-19 № 217
на № _____ від _____

АКТ

про впровадження результатів дисертаційної роботи Путренка Віктора Валентиновича на тему «Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку» поданої на здобуття наукового ступеня доктора технічних наук за спеціальністю 01.05.04 – Системний аналіз і теорія оптимальних рішень

Навчально-науковий комплекс «Світовий центр даних з геоінформатики та сталого розвитку» КПІ ім. Ігоря Сікорського веде багаторічну дослідну роботу в напрямі математичного моделювання та оцінювання процесу сталого розвитку складних соціально-еколого-економічних систем на основі використання великих масивів даних. В цьому контексті робота Путренко В.В. є дуже актуальною, оскільки в ній проведено аналіз методології інтелектуального аналізу геопросторових даних з метою розпізнавання прихованих закономірностей та відношень в задачах сталого розвитку територіально розподілених систем.

Результати дисертаційної роботи Путренка Віктора Валентиновича на тему «Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку» впроваджено у наступних науково-дослідних та прикладних темах, які виконувались в ННК СЦД-України:

- Розробка методології та математичної бази кількісної оцінки процесів сталого розвитку та впливу на них сукупності загроз як у глобальному, так і в регіональному контексті (0111U001738)
- Форсайт економіки України: середньостроковий (2015–2020 роки) і довгостроковий (2020–2030 роки) часові горизонти
- Міжнародний проект EU funded Programme for the prevention, preparedness, and response to man-made and natural disasters in the ENPI East Region – EuropeAid/129397/C/SER/Multi (EP 105044)
- Аналіз соціально-економічних процесів розвитку суспільства, заснованого на знаннях (0114U000670)
- Розробка он-лайн платформи аналізу і сценарного планування сталого розвитку регіонів України в контексті якості та безпеки життя людей (0117U002476)
- Міжнародний проект 6166 спільно з УНТЦ та НАН України «Розробка інформаційних технологій для моделювання, кількісного оцінювання та прогнозування впливу загроз виникнення конфліктів та розповсюдження зброї на стабільність розвитку суспільства в регіональному та глобальному масштабах» (STCU #6166)

Використання великого обсягу емпіричних даних під час дослідницької роботи дозволило Путренко В.В. обґрунтувати методологію інтелектуального аналізу геопросторових даних для аналізу процесу сталого розвитку на основі методів класифікації, кластеризації, геостатистичного аналізу, нейромережових методів з використанням даних дистанційного зондування Землі. Отримані в ході виконання дослідницьких робіт результати представлено у вигляді звітів, картографічних матеріалів, інтерактивного інструментарію на базі ГІС, методичних рекомендацій, публікацій у міжнародних та національних наукових виданнях, впроваджено у моделюванні сценаріїв розвитку складних систем різної природи в ситуаційному центрі.

Директор
ННК «Світовий центр даних
з геоінформатики та сталого розвитку»
КПІ ім. Ігоря Сікорського



К.В. Єфремов



НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ І ГЛОБАЛЬНОГО ІНФОРМАЦІЙНОГО ПРОСТОРУ

Чоколівський бул. 13, м.Київ, 03186, тел/факс (044) 245-88-38, тел. 245-87-97
 E-mail: itelua@kv.ukrtel.net P/p 35219001000323 в УДК м. Києва, ЗКПО: 26022051, МФО: 820019

12.11.19 № 167/12.11.19-2 На № _____ від _____

ЗАТВЕРДЖУЮ

Директор інституту

Д.т.н., проф., чл.-кор. НАН України
 Трофимчук О.М.



[Handwritten signature]

про впровадження результатів дисертаційної роботи **Путренка Віктора Валентиновича** на тему «**Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку**» поданої на здобуття наукового ступеня доктора технічних наук за спеціальністю 01.05.04 – Системний аналіз і теорія оптимальних рішень

В Інституті телекомунікацій і глобального інформаційного простору Національної академії наук України розглянуто методологію інтелектуального аналізу геопросторових даних для використання в системах прийняття рішень в задачах дослідження процесів сталого розвитку регіонів України в контексті підвищення якості та безпеки життя людей.

Методологія призначена для оцінювання параметрів сталого розвитку з використанням геопросторових даних на основі системного підходу. Запропонована концепція класифікації та обґрунтовано доцільність використання інтелектуальних методів аналізу геопросторових даних для сценарного моделювання сталого розвитку на засадах методів системного аналізу. Основною метою використання інтелектуального аналізу геопросторових даних є пошук закономірностей та відносин у великих масивах даних, які містять просторово координовану прив'язку. До основних напрямів ІАГД відносяться класифікація, кластеризація, правила асоціацій, геостатистика та геовізуалізація, які разом формують методологію системного інтелектуального аналізу з метою підтримки прийняття рішень. Базові методи інтелектуального аналізу геопросторових даних визначаються типом

розподілу даних та гіпотезою імовірності оцінки виникнення аномальних значень на обмеженій просторовій відстані. Доведено доцільність застосування трендових поверхонь для визначення глобальних трендів регіональної зміни параметрів сталого розвитку

Інститут телекомунікації і глобального інформаційного простору НАН України повідомляє про впровадження результатів дисертаційної роботи Путренка Віктора Валентиновича на тему «Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку» в практичній роботі та при моделюванні індикаторів сталого розвитку на основі даних дистанційного зондування Землі та геопросторової інформації.

Завідуючий відділом, д.ф.-м. н.



Полумієнко С.К.

Заступник директора з наукової роботи, к.т.н.



Лебідь О.Г.

Учений секретар, к.т.н., с.н.с.



Клименко В.І.

МІНІСТЕРСТВО РЕГІОНАЛЬНОГО РОЗВИТКУ,
БУДІВНИЦТВА ТА ЖИТАОВО-КОМУНАЛЬНОГО ГОСПОДАРСТВА УКРАЇНИ

ДЕРЖАВНЕ ПІДПРИЄМСТВО
УКРАЇНСЬКИЙ ДЕРЖАВНИЙ НАУКОВО-ДОСЛІДНИЙ
ІНСТИТУТ ПРОЕКТУВАННЯ МІСТ

ДІПРОМІСТО
ІМЕНІ Ю.М. БІЛОКОНЯ

бул.Л.Українки, 26
м.Київ
01133
Україна

+38 044 285 6338
+38 044 285 6372 факс
admin@dipromisto.gov.ua
www.dipromisto.gov.ua



Розрахунковий рахунок № 26008319423
в АТ "Райффайзен Банк АВАЛЬ" в м. Києві
МФО 380805
Код ЄДРПОУ 02497720

№ Гіс-2493 від 13.11.2019
на № _____ від _____

Національний технічний університет України
«Київський політехнічний інститут
імені Ігоря Сікорського»

АКТ

**про впровадження результатів дисертаційної роботи Путренка Віктора Валентиновича
на тему «Методологія інтелектуального аналізу геопросторових даних для задач
сталого розвитку» поданої на здобуття наукового ступеня доктора технічних наук за
спеціальністю 01.05.04 – Системний аналіз і теорія оптимальних рішень**

В Державному підприємстві "Український державний науково-дослідний інститут проектування міст "ДІПРОМІСТО" імені Ю.М. Білоконя розглянуто методологію інтелектуального аналізу геопросторових даних для використання в системах прийняття рішень в задачах дослідження процесів сталого розвитку регіонів України в контексті підвищення якості та безпеки життя людей.

Запропонована автором концепція класифікації та обґрунтування доцільності використання інтелектуальних методів аналізу геопросторових даних для моделювання сталого розвитку на засадах методів системного аналізу є актуальним та перспективним напрямом досліджень, який було інтегровано до задач територіального планування на регіональному та місцевому рівні створення планувальної документації та розробки стратегії розвитку території.

Основною метою використання інтелектуального аналізу геопросторових даних є пошук закономірностей та відносин у великих масивах даних, які містять просторово координовану прив'язку. Важливою складовою інтелектуального аналізу геопросторових даних є моделювання просторових відносин методами зворотної відстані, діапазони відстаней, зон індиферентності, суміжності та сусідства. Запропонований підхід до застосування нейронних мереж дозволяє оцінити параметри та інтегральну оцінку складових сталого розвитку на різних територіальних рівнях, які відповідають європейській класифікації регіонів NUTS 1-5 та державного адміністративно-територіального устрою України, включаючи новий тип адміністративних територіальних утворень – об'єднані територіальні громади.

Державне підприємство "Український державний науково-дослідний інститут проектування міст "ДІПРОМІСТО" імені Ю.М. Білоконя повідомляє про впровадження результатів дисертаційної роботи Путренка Віктора Валентиновича на тему «Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку» в практичній роботі та під час створення планувальної та містобудівної документації з використанням запропонованих методів аналізу геопросторових даних та підготовки відповідної картографічної продукції засобами геоінформаційних систем.

Заступник директора з наукової роботи,
керівник Центру ГІС ДП "ДІПРОМІСТО",
доктор географічних наук, проф.



Ю.М. Палеха



УКРАЇНА

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

03056, м. Київ, пр-т Перемоги, 37; тел. (+38 044) 204-82-82 тел./факс (+38 044) 204-97-88
<http://www.kpi.ua> e-mail: mail@kpi.ua ЄДРПОУ 02070921

11.11.2019 № 33

Довідка

про впровадження результатів
докторської дисертаційної роботи
Путренка Віктора Валентиновича

Довідка видана Путренку Віктору Валентиновичу у тому, що результати виконання його дисертаційної роботи «Методологія інтелектуального аналізу геопросторових даних для задач сталого розвитку», зокрема:

- метод моделювання на основі географічно зваженої регресії та локального індексу Морана
- метод розрахунку індексу економічного розвитку території з використанням нейронної мережі на основі навчання на базі даних дистанційного зондування Землі
- методика трендового аналізу функціонального та кількісного зв'язку вкладу складових індексу сталого розвитку
- методологія використання інтелектуальних методів аналізу геопросторових даних
- методи геостатистичного аналізу та оптимізації задачі інтерполяції даних стохастичними методами крігінгу
- методи просторової кластеризації з використанням статистичної обробки «гарячих-холодних» точок

використовуються на кафедрі математичних методів системного аналізу при викладанні Путренко В.В. таких курсів:

1. Основи геоінформаційних систем
2. Інтелектуальний аналіз даних в геоінформаційних системах

Заступник директора ІПСА КПІ ім. Ігоря Сікорського
Д.т.н., професор

Професор ІПСА КПІ ім. Ігоря Сікорського, д.т.н.



В.Д. Романенко

П.І. Бідюк